HG1710
E9

WORKING PAPER ONLY - DO NOT QUOTE
DOCUMENT DE TRAVAIL SEULEMENT - NE PAS CITER

Electronic Funds Transfer (EFT) and the Industrial Organization of the Canadian Banking Industry

A Report Prepared for

Department of Consumer and Corporate Affairs
Government of Canada

by

B. Curtis Eaton and

G. Frank Mathewson
(Mathewson Economic Consultants Inc.)

L 1981

The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the Department of Consumer and Corporate Affairs.

Contents

		Page
Chapter 1	A Model of Banking Services and EFTS	1
Chapter 2	Selected Facts on the Canadian Banking Industry and a Selected Survey of Previous Work on the Economics of Banking	24
Chapter 3	Economic Features of EFT: Competition and Cooperation	39
Chapter 4	A Summary of the Principal Results	56

Preface

This report summarizes our inquiry into the economic impact of EFT on the structure, conduct and performance of the Canadian banking industry. We have surveyed on extensive literature on (i) strategic behavior in oligopolies (ii) nature of the Canadian banking industry including the impact of competition from near-banks (iii) the technical and economic characteristics of EFT and (iv) the impact of government regulation and legislation on the Canadian banking industry.

Chapter 1 sets forth the essential ingredients of an economic model that specifies the nature of the banking service, identifies the features of that service relevant to EFT, and analyzes the impact of EFT on the equilibrium features of the industry. Chapter 2 reviews selectively the economic literature on banks relevant to this study; as well, we evaluate the empirical methodology specific to much of this literature. Chapter 3 evaluates the empirical economic literature on EFT including two illustrations of the possible impact of EFT on Canadian banking. Finally, Chapter 4 summarizes our results and includes several conjectures and issues for future research.

Chapter 1 A Model of Banking Services and EFTS

1.1 The Specialized Services of Commercial Banks

The receipt of earnings by individuals over their lives follows a typical and well-known pattern. In their formative years, earnings are typically low or non-existent because the individual specializes in the acquisition of human capital. In the early working years (from roughly age 20 to age 40), the individual continues to acquire human capital and earnings are typically positive and increasing with age. In the middle years of the individual's working life (from roughly age 40 to age 55) earnings increase slowly (or not at all) with age and in the last 10 years of working life, earnings remain constant or decrease with age. After age 65 earnings typically decrease radically and are frequently non-existent. Over this entire life cycle, expenditures on goods and services are roughly equivalent to earnings, but the typical pattern of expenditure over the life cycle does not closely match the pattern of earnings. Rather, expenditures exhibit less fluctuation. They exceed earnings in the early working years and in the post-retirement years and fall short of earnings in the middle years. The whole range of financial intermediaries are, to some extent, useful to the individual in his attempt to "solve" the life-cycle problem.

The individual faces a similar problem over much shorter periods of time. Typically, the receipt of liquid assets (earnings or borrowed funds, for example) is, relative to expenditure, lumpy and infrequent. Thus, over short periods of time (a month, or a year for example), the receipt of liquid assets and expenditure from them are poorly synchronized. Receipts are lumpy

and infrequent while expenditures are continuous. From the individual's perspective, the services which commercial banks provide are particularly useful in the endeavour to manage this short-term "working-balance problem". In particular, demand deposits and bank credit cards are ideal instruments for this purpose. From the individual's point of view, commercial banks can be regarded as intermediaries which provide services which are specialized to the management of short-term working balances.

This view, of course, corresponds to the traditional distinction between commercial banks and other financial intermediaries. Bond and Shearer (1972, p. 168), for example, observe that "a pure commercial bank would have short-term, highly volatile deposit liabilities which would be used as money in the normal course of events." Near banks are conceptually defined as institutions which raise a substantial proportion of their funds by accepting deposits which are close substitutes for personal chequing and savings deposits of commercial banks. Although the relative sizes of commercial banks and near banks has changed significantly over the past few decades, it still remains true that the direct involvement of near banks in the payments mechanism is not substantial.

The inducements to commercial banks to provide such services are obvious--service fees, the opportunity to lend idle balances and the opportunity to earn interest on consumer credit extended. The inducements to the individual are <u>security</u>, <u>liquidity</u> and the opportunity to <u>earn interest on idle balances</u>.

1.2 <u>Resource Costs Born by Banks and their Customers</u>

A striking and, for current purposes, fundamental property of this market for banking services is that the real resources (primarily time) devoted by individuals to acquire these services are substantially larger than the real resources used by commercial banks to provide them.

. . . . In many of its aspects the bank is a manufacturing firm: acquiring funds, processing pieces of paper, manipulating data and producing loans, investments and deposit services. But a characteristic of banks that distinguishes them from, say, auto manufacturers, is the fact that consumption of a bank's services requires inputs of the consumer's time and energy which are large relative to the bank's own costs. When the customer writes and mails a check, the banking system accepts the check from the payee, moves it through a national distribution system to the payor's bank, processes it and returns the funds to the payee and the check to its depositor. The costs to the bank of this activity are important to it, but the customer has his own costs: postage and stationery to mail the check and a variety of time costs, such as time to write the check, to do his own accounting to reconcile balances when the check is returned. In all this, the customer's costs for a check typically exceed those of the bank. In the same way, a customer making a deposit at a bank imposes accounting and handling costs on that bank. But his own costs of going to the bank, waiting in line and returning to his normal activity add up, at reasonable opportunity costs for his time, to more than those of the bank.

Baxter, Cootner and Scott (1977, pp. 16-17).

The value of a minute to an individual who earns $20,000 a year is $.18. If a transaction (deposit or withdrawal, for example) requires 30 minutes of the individual's time (20 minutes to get to the bank and return, and 10 minutes in the bank), then the individual has contributed more than $5.00 of time. The real resources which the bank devotes to such a transaction are estimated (Walker 1979) to be less than $1.00. Writing cheques and verifying bank balances is similarly intensive in the individual's time--10 minutes per cheque is not an unreasonable estimate.

For current purposes these elementary observations are fundamentally important. The management of working balances requires a large number of transactions, all of which impose significant direct costs on the individual. Various forms of EFT offer the possibility of a substantial reduction in these transactions costs.

1.3 Direct versus Indirect Access to Bank Services

Thus the services which commercial banks provide individuals involve costs which are borne directly by the individual. The costs which the individual actually bears are, of course, controlled by the individual. Individuals drawn down working balances or increase consumer debt in order to settle accounts with third parties (retailers in particular). It is important to distinguish two means by which individuals access the services of commercial banks. The writing of a cheque in favour of some third party is a form of direct access since the individual settles the account directly with the third party and no intermediate transaction between the bank and the individual is required. A cash settlement, on the other hand, is a form of indirect access--the individual must first withdraw cash from the bank and only then

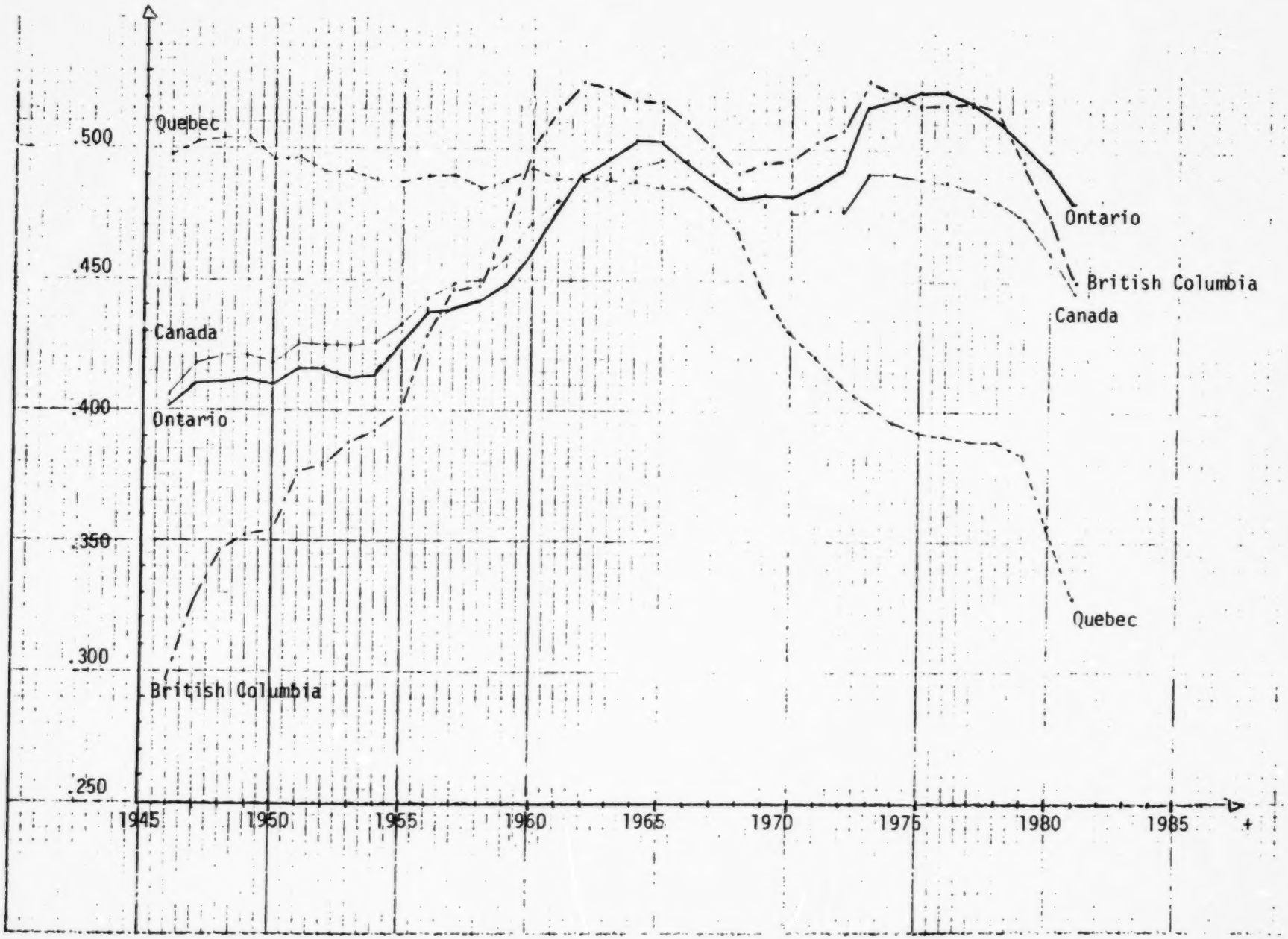
can he/she settle accounts with the third party. Some accounts can be settled only with cash--vending machines require cash; retailers often refuse to accept cheques to protect themselves from fraud; many individual's prefer cash payments for tax reasons. Among those accounts which can be settled with a cheque or cash, the individual has the choice of means of payment. This choice is undoubtedly conditioned by the costs to the individual of direct access (writing the cheque, waiting for verification, checking the balance at the end of the month) relative to indirect access (going to the bank, waiting in line). Elementary economic considerations suggest that the individual will choose to settle small accounts with cash and large accounts by cheque.

It is also useful to distinguish between direct and indirect access to consumer credit. The use of a bank credit card is a form of direct access to credit, whereas a trip to the bank to negotiate a loan in order to settle accounts with some third party is a form of indirect access.

1.4 Instruments of Competition in Commercial Banking

Commercial banks compete with each other for the deposits from and loans to individuals through a variety of price and non-price instruments. Interest and service fees are the primary price instruments. An obvious and important non-price instrument of competition is the network of branches maintained by commercial banks. (Figure 1.1, plots branches per capita from 1946-1983 in Canada, British Columbia, Ontario and Quebec.) With respect to non-price instruments, it is useful to distinguish those which affect the individual's costs of indirect accessibility

Figure 1.1


branches/thousand pop ≥ 20
.500
.450
.400
.350
.300
.250
Quebec
Canada
Ontario
British Columbia
1945
1950
1955
1960
1965
1970
1975
1980
1985
+
Ontario
British Columbia
Canada
Quebec

from those which affect the costs and scope of direct accessibility. The branching network is obviously an important determinant of the costs of indirect accessibility--the denser is the branching network, the closer are individual customers to their local branch and the lower are their time costs of indirect accessibility. Chequeable accounts and bank credit cards are an important determinant of the scope of direct accessibility.

The introduction of on-line debit-credit cards (point-of-sale transfer of funds) would, in the opinion of most observers, significantly reduce the costs to the individual of direct accessibility. In particular, it would eliminate the time required for cheque and credit verification, and may also simplify the individual's "bookkeeping" problem. In addition, it could increase the scope of direct accessibility. The critical factor is the response of retailers to EFTS. In accepting an individual's personal cheque, the retailer incurs a "bad-cheque" risk and, in addition, is effectively extending credit to the purchaser for the period of time required for the cheque to clear. Point-of-sale transfer eliminates both of these costs, a clear inducement for retailers to be more receptive to point-of-sale transfer than they are to accepting personal cheques. Further, point-of-sale transfer should reduce the time cost to the retailer of direct accessibility by eliminating the time required to verify cheques. Against this, one must weigh the costs to the retailer of tieing into the transfer network. Others (Flannery et al. 1975, Baxter et al. 1977) argue that on balance there is an incentive for most retailers to use EFTS. Point-of-sale transfer thus promises to decrease the costs to the individual of direct accessibility and to increase the scope of direct accessibility. A significant factor of point-of-sale transfer is that it inevitably must involve the electronic transfer funds from accounts in one bank to accounts in other banks. Point-of-sale transfer is a form of inter-bank electronic transfer.

A second form of electronic transfer is intra-bank electronic transfer. The cash dispensers/automated tellers (which we label cash-vending for convenience) which are already being introduced in Canada and elsewhere are illustrative of intra-bank electronic transfer. Cash-vending potentially reduces the cost to the individual of indirect accessibility in three ways. It should reduce the time spent queueing in the bank; it effectively increases banking hours for purposes of indirect accessibility; finally and most significantly, it increases the number of locations where indirect accessibility is possible. Cash-vending then decreases the individual's costs of indirect accessibility. From the individual consumer's perspective, inter-bank transfer changes the cost and potential scope of direct access while intra-bank transfer changes the cost of indirect access.

1.5 The Prospects for and Implications of EFT: A Conceptual Framework

There are two natural questions to pose initially on EFTS. First, what are the incentives of individual commercial banks or commercial banks collectively to introduce some form of EFT? We focus on two forms: intra-bank cash-vending and point-of-sale transfer. The relevant incentives for intra-bank transfers are the individual bank's; the relevant incentives for point-of-sale transfers are the collective incentives of all commercial banks. Second, what are the implications for individuals, for commercial banks and for retailers of various forms of EFT? Commercial banks are the "prime movers" in this scenario as commercial banks will choose whether to introduce these innovations. Therefore, in assessing the prospects for any form of innovation, the crucial issue is the implication for bank profits of any innovation.

The economic response of individual depositors and borrowers to an innovation is the first issue which must be addressed. From the bank's point of view, a clear understanding of this response is obviously necessary for any

evaluation of the potential profitability of an innovation. We address this problem in Section 1.6 below. The second issue which must be addressed is the competitive response within the commercial banking industry to any innovation. The introduction of cash-vending and/or point-of-sale transfer clearly changes the economic environment in which banks compete, changing the degree and perhaps even the nature of inter-bank competition. For example, what are the implications of increased direct accessibility on the branching strategy and pricing behaviour of banks? Again from the commercial bank's perspective, a clear understanding of the competitive responses within the industry is necessary for a rational evaluation of the potential profitability of an innovation. We address these issues sequentially. In Section 1.7 we consider the price response to innovations, holding the branching network fixed. In Section 1.8, we consider responses in branching strategy, the potential profitability of innovations and their implications for individual customers.

Our conception of the problem draws heavily on the recent literature on strategic behaviour in oligopolies. We append to this report a survey of the relevant literature with some non-banking illustrations. (The Appendix is drawn from our progress report of March 30, 1983.) In the terminology of that survey, we view branches and the capital costs associated with EFT as precommitted variables and deposit rates as non-committed variables.

1.6 Accessibility-Security-Interest: The Individual's Problem

To analyze the individual's management of working balances we adapt a simple, but insightful model due to Baumol (1952) and Tobin (1956). We consider the management of working balances for an individual who is drawing on lumpy and infrequent income receipts to finance continuous consumption expenditures. A similar problem arises for individuals who are drawing down a credit line

and/or negotiating lumpy consumer loans. Since our primary concern at this state is to illustrate the type of modelling which is required to assess adequately the prospects for and implications of various forms of EFT we ignore these credit line/consumer loans problems to keep the model relatively simple and interpretable.

The individual in this model receives a lumpy wage payment at the beginning of each period. The payment is equal to W and the individual wishes to consume continuously over the period. For simplicity we suppose that W is deposited in the individual's demand account in some commercial bank which pays interest (on the average balance held over the period) at rate r. (To keep the model simple, we interpret r as a rate net of service fees.) Indirect accessibility, or cash withdrawal, is costly to the individual since it involves a trip to the bank. We assume that the total costs to the individual of each transaction with the bank have a dollar equivalent equal to c. The aggregate costs of n transactions are nc.

We assume that a proportion α of all transactions can be settled via direct access to deposit balances (by writing a cheque, for example) and a residual proportion $(1-\alpha)$ require cash and hence indirect accessibility. We assume that where feasible, direct access is preferred to indirect access. Thus the individual will use direct access to spend at rate αW and will use cash to spend at rate $(1-\alpha)W$ over the period.

Given n, the timing and magnitude of cash withdrawals which maximizes interest income, consistent with the desire to spend cash at rate $(1-\alpha)W$, is to choose evenly-spaced withdrawals of magnitude $(1-\alpha)W/n$--the first withdrawal will occur at time 0 (the beginning of the period) and the last at time $1-1/n$. The average balance held in $\alpha W/2$ plus $(1-\alpha)W(n-1)/2n$. The average balance and hence interest income is an increasing function of n, but indirect accessibility is costly. Thus the individual faces a trade-off between interest income on working balances and the costs of indirect accessibility.

We capture the individual's concern for security by supposing that funds on deposit are perfectly secure and that the probability of loss or theft of cash is ρ (per period). Given n, the average cash holding is $[(1-\alpha)W]/[2n]$ and thus the "expected loss" is simply $[\rho(1-\alpha)W]/[2n]$. We assume that the individual is risk neutral. The management-of-working-balances problem this reduces to:

$$\text{Max } U = r\left[\alpha W/2 + (1-\alpha)W(n-1)/2n\right] - \rho(1-\alpha)W/2n - nc. \tag{1}$$

The first term is interest income earned on the average working balance; the second is the expected loss associated with the insecurity of holding cash; and the third is the cost of indirect accessibility. The solution to this problem yields:

$$n^* = \left(\frac{(1-\alpha)W(n+\rho)}{2c}\right)^{\frac{1}{2}}, \tag{2}$$

and, in turn,

$$u^* = \frac{rW}{2} - \left(2(1-\alpha)Wc(n+\rho)\right)^{\frac{1}{2}}. \tag{3}$$

The average deposit held over the period is:

$$D^* = \frac{W}{2} - \left(\frac{(1-\alpha)Wc}{2(r+\rho)}\right)^{\frac{1}{2}} \tag{4}$$

Thus the number of transactions and the average deposit level are directly related to W, r and ρ and inversely related to c. Further, direct accessibility is clearly valuable to the individual (i.e., U^* is an increasing function of α) since direct accessibility allows the individual to economize on resources devoted to indirect access.

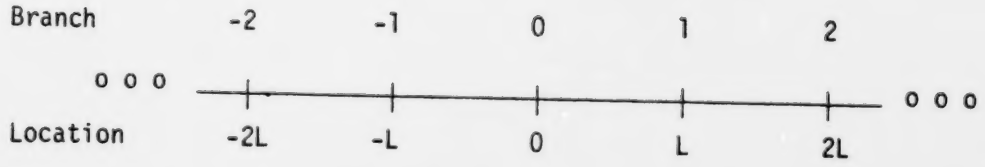
We are not yet in a position to consider carefully the issues raised by various forms of EFT, but we can mention some clear non-sequiturs which abound in the EFT literature. Many observers argue that point-of-sale transfer will substantially enhance direct accessibility, since, they argue, it will reduce transactions costs borne by retailers and by individuals. That is, more retailers will be receptive to point-of-sale transfer than to other forms of direct access (cheques in particular) and their customers will also find point-of-sale transfer attractive. Further, point-of-sale transfer is thought to be cost saving to commercial banks as well. Thus, it appears to be attractive to all concerned parties. In particular, given r and c, it is attractive to commercial banks since it decreases their costs and increases the supply of deposits to them (see equation (4)). Thus, it is only a matter of time until we observe this fundamental change in the payments mechanism. The difficulty with this argument is that r and c cannot be taken as given and hence it is not clear that banks, the prime movers, have the incentive to introduce this innovation. Indeed, as we see below it is not at all obvious that bank's have the incentive even if the innovation would increase the supply of deposits and reduce their costs.

1.7 Price Response to Innovations in the Payments Mechanism

In this section we develop a simple model of spatial competition among banks for the deposits of individuals. Our objective is to determine the implications of point-of-sale transfers and cash-vending on the supply of deposits to commercial banks and hence on their pricing behaviour, given both the branching network and the innovations themselves. We discuss implications of these innovations for the branching network and the banks' incentives to introduce them in Section 1.8 below.

Figure 1.2

Location of Bank Branches



For concreteness and ease of illustration, we imagine a one-dimensional physical market (a road, for example) along which identical individual households are uniformly and continuously distributed. We suppose also that branches of commercial banks are evenly spaced along the one-dimensional market, at discrete intervals of length L.

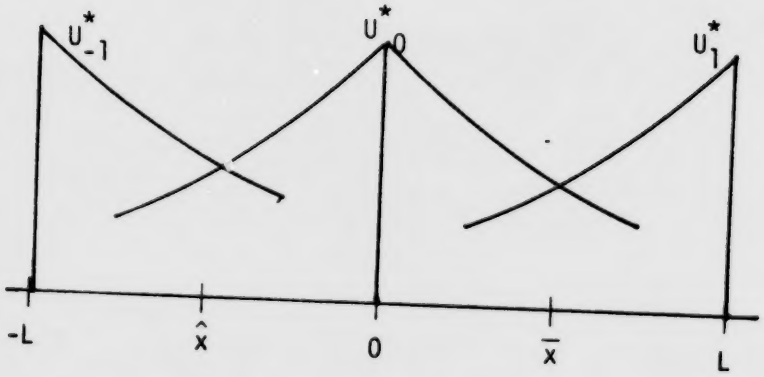
Figure 1.2 is an illustration of the location of branches. We choose the location of some branch, branch 0, as an arbitrary origin and number branches to the right by 1, 2, ... and those to the left by -1, -2, ... We assume, for convenience, and without great injury to the facts, that no two adjacent branches are owned by the same commercial bank.

We begin by examining the supply of deposits to a representative branch, branch 0 in Figure 1.2. Branch 0 will be in competition with branches 1 and -1 for the deposits from individuals located in the interval -L to L. Let r_0, r_1, and r_{-1} denote the deposit rates offered by branches 0, 1 and -1 respectively. The distance of the individual from a branch determines the cost of indirect access to that individual from that branch. We assume that the cost of indirect accessibility is an increasing function of distance, $c = C(\cdot)$. Measuring distance from the origin in Figure 1.2 by x, the cost of indirect access to branch 0 for an individual located at x is $C(x)$, and the cost of indirect access to branch 1 is $C(L-x)$ since $L-x$ is the distance from branch 1.

Individuals in interval (0,L) will choose to patronize either branch 0 or branch 1. For x in this interval, maximized utility (equation (3)) associated with branch 0 decreases with x, and maximized utility associated with branch 1 increases with x--the observations follow from the fact that the costs of indirect access increase with distance. In Figure 1.3 U_0^* denotes

Figure 1.3

Consumer Choice of Branch



maximized utility of patronizing branch zero and U_1^* is maximized utility of patronizing branch 1. Individuals located in interval $(0,\bar{x})$ will choose to patronize branch 0 and individuals in $(\bar{x},L)$ will choose to patronize branch 1. Similarly, individuals in interval $(\hat{x},0)$ will choose to patronize branch 0 (and those in $(-L,\hat{x})$ to patronize branch -1). The "market boundary" $\bar{x}$ is determined by r_0 and r_1 (among other things) and $\hat{x}$ is determined by r_0 and r_{-1}: we write $\bar{x}(r_0,r_1)$ and $\hat{x}(r_0,r_{-1})$. From equation (4) we observe that the supply of deposits from an individual will be determined by the rate of interest available on a deposit and by the distance from the branch, since distance determines the cost of indirect accessibility. We write $D^*(r_0,|x|)$ to denote the supply of deposits from an individual to branch 0 since $|x|$ is the distance from the branch.

Thus, the aggregate supply of deposits to branch zero is

$$S^0(r_0,r_1,r_{-1}) = \int_{\hat{x}(r_0,r_{-1})}^{\bar{x}(r_0,r_1)} D^*(r_0,|x|)dx \qquad (5)$$

The most important feature of this supply function is that it is an increasing function of r_0: that is, the branch has monopsony power. As r_0 increases, the supply of deposits to branch zero increases on two accounts: First, from equation (4), we see that an increased deposit rate induces each individual to increase his/her average balance; second, as r_0 increase $\bar{x}$ moves to the right (and $\hat{x}$ to the left) which reflects the fact that the number of individuals who choose to patronize branch zero increases with r_0.

Let us suppose that commercial banks set deposit rates in a non-cooperative fashion (i.e. they are <u>not</u> a cartel). This assumption is not crucial to our conception of the problem (as we will see below) but it is

convenient for purposes of illustration. If we assume further that the markets for the instruments which banks hold as assets (loans, securities, etc.) are competitive, we can express the excess of revenues over costs for branch 0 (the "gross profits" of branch 0) as

$$G(r_0,r_1,r_{-1}) = (i-r_0-z)\ S^0(r_0,r_1,r_{-1}) \tag{6}$$

where i is the rate of return on assets and z is the variable cost per dollar to the branch of holding deposits. (We assume that costs of holding assets are negligible.) The choice of r_0 which will maximize $G(r_0,r_1,r_{-1})$, r_0^*, will satisfy the standard first order condition:

$$-\ S^0(r_0^*,r_1,r_{-1}) + (i-r_0^*-z)\ S_1^0(r_0^*,r_1,r_{-1}) = 0 \tag{7}$$

where $S_1^0 \equiv \partial S^0/\partial r_0^*$.

Given our assumptions that no commercial bank owns two adjacent branchs and that all branches are evenly spaced, we can characterize the non-cooperative price equilibrium in this market by

$$-\ S^0(r^*,r^*,r^*) + (i-r^*-z)\ S_1^0(r^*,r^*,r^*) = 0 \tag{8}$$

where r^* is the equilibrium rate of interest on deposits. Clearly, $r^* < i-z$ and the gross profits of commercial banks will be positive. This reflects the monopsony power which is inherent in this market.

We can now consider several issues raised by the development of EFT. Let us consider point-of-sale transfer first. There are two features of point-of-sale transfer which deserve consideration; we consider them separately. The parameter z captures the costs to the bank of handling

deposits. To the extent that point-of-sale transfer decreases these costs, z will decrease. This effect captures the substitution of a less costly form of direct access for the current form (cheques). Economics then reveals that r^* will increase as z decreases as will maximized profits of commercial banks. On this score, and holding the branching network fixed, banks and their customers will share the rent attributable to decreased costs.

In addition, as we have argued, point-of-sale transfer is likely to increase direct accessibility as well. Here the story is somewhat more interesting. The crucial response is the response of r^* to increased accessibility. We have captured the degree of direct accessibility by α; thus we are interested in the comparative static result: $dr^*/d\alpha$. In general, further information is required to sign this derivative. However, one result is easily verified: As α approaches 1 (as direct accessibility completely replaces indirect accessibility), $dr^*/d\alpha$ is positive and r^* approaches $i-z$. That is, as α goes to 1, the deposit rate approaches the "competitive" rate and the gross profits of commercial banks are eliminated. In this case, all of the rent created by point-of-sale transfer accrues to depositors. Thus, we may hazard a conjecture: To the extent that point-of-sale transfer increases direct accessibility we expect, holding the branching network fixed, increased deposit rates and decreased gross profits for commercial banks.

The intuition of this asymptotic result is straightforward and appealing. The existence of positive gross profits for commercial banks is driven by market power, by the spatial monopsony power inherent in the supply of deposits to individual branches. But, this monopsony power is derived in turn solely from indirect accessibility: Proximity to a branch is valuable to the depositor only to the extent that a depositor needs to conduct business through that

branch. As α goes to 1, a depositor has no occasion to visit the branch and proximity to it has no value to a depositor. In this limit, the choice of branch is determined exclusively by deposit rates and a depositor will choose the branch offering the highest deposit rate. The monopsony power is completely eroded in the "cashless" society.

We now turn to the price implications of cash-vending. We argued above that cash-vending served to decrease the costs to individuals of indirect accessibility. In the context of our model of spatial competition, we can capture cash-vending by a reduction in transport costs. Assuming that these costs are linear in distance yields

$$C(d) = td \tag{9}$$

We can approximate an increase in the extent of cash-vending by a decrease in t. (This is only a proxy since our model does not capture the use by individuals of cash-vending services at more than one location in the market space.) The comparative static dr^*/dt is negative. This implies that the introduction of cash-vending tends to increase deposit rates, holding the branching network fixed. Maximized bank profits may increase or decrease, but there is a presumption that they will decrease since, as t approaches zero, we again converge to the competitive limit.

1.8 The Branching, Network and the Prospects for EFTS: Conjectures and Speculations

A model of the banking industry equilibrium is essential for an understanding of the implications of innovations in the payments mechanism for the existing network of bank branches. For each bank, individual branches once established represent a substantial sunk cost. Branches once established, therefore, cannot be easily or costlessly relocated. This feature has a strategic advantage. It means that branches can be used strategically to deter subsequent entry into the industry: Establish a branch to divide up the market so that residual markets will not sustain the branch of a competing bank. Other banks knowing the sunk cost of a branch at a specific location know that even if they enter the branches of the existing banks will not be relocated. Analyses of strategic behaviour of this type appear in the economics literature by Schmalensee (1978), Eaton-Lipsey (1979) and Eaton-Wooders (1983). Following this approach, we may define a banking industry equilibrium as the establishment of the branching network which maximizes the profits of the industry as a whole, subject to the condition that entry is unprofitable. This view ignores some important problems having to do with rivalry between existing firms within the industry. (e.g., The monopsony power implicit in this argument may be eroded by existing competition.) These problems of limited monopolistic competition have not been modelled in a strategic spatial context in the theoretical literature but some arguments suggest that this spatial pre-emptive view of an industry equilibrium with some non-cooperative firm competition is plausible. (For example, see the discussion in Section VI.b the survey article by Archibald-Eaton-Lipsey (forthcoming).)

If such arguments capture the underlying equilibrium of bank competition, we may consider the implications for the branching network of innovations in the payments mechanism. Consider first point-of-sale transfer and, as before, distinguish those implications arising from the substitution of point-of-sale transfer for inferior forms of direct access from those implications arising from increased direct access. What are the various forces at work from the introduction of EFTS?

To the extent that point-of-sale transfer simply substitutes for cheques, we expect an increase in the density of the network of branches. This substitution increases the profits of commercial banks and the profits of potential entrants (assuming that these entrants have access to the EFT network). An increased density of branches is required to re-establish an industry equilibrium.

To the extent that point-of-sale transfer increases the scope of direct access, the implications are different. A substantial increase in direct accessibility yields a decrease in the profits of commercial banks as the spatial monopsony power of branches is eroded and a consequent decrease in the density of the branching network. Insight emerges if we consider the limiting case in which direct accessibility completely displaces indirect accessibility. Then, the branching network should largely disappear.

Do <u>individual</u> commercial banks have a long-run incentive to introduce point-of sale transfer? In our view, the answer is no. <u>Collective</u> action by all banks is required. The substantial economies of scale plus the network features of point-of-sale transfer mean that such innovation is prohibitive for individual banks. Whether banks collectively have such an action depends upon the economic terms to access the network for potential entrants.

Suppose first that entrants can access the network at marginal cost. Then any innovative bank(s) would be instituting a network and selling access at a price below average cost to its (their) competitors. This would only enhance competition for the banks eroding any existing monopsony power. This is particularly true if EFTS means a substantial increase in direct accessibility; enhanced direct accessibility tends to destroy the monopsony power inherent in the industry.

The prospects are substantially different if banks can control the terms of access to entrants. The ownership and control of the network then creates a natural monopoly, not unlike the natural monopoly (currently under erosion in the U.S.) in the telephone industry. Further discussion of this natural monopoly element appears in Chapters 3 and 4.

We have already seen the introduction of some equipment such as cash-vending and other forms of intra-bank electronic transfer in commercial banks in Canada. An important distinction between the intra-bank electronic transfer network and point-of-sale transfer is that existing automated practices do not require extensive inter-firm cooperation. In this case, individual banks can (and indeed have) initiate the innovation. Our analysis predicts that individually initiated forms of electronic transfer can enhance price competition among commercial banks; in a long-run industry equilibrium we cannot predict whether such innovation will increase bank profits, whatever the effect on any individual bank's short-run profits. In either case, such innovation will yield a decrease in the density of the network of branches as they exist. The reason is simple: Cash-vending within the commercial banking industry tends to increase the scale at which potential entrants must operate. The introduction of cash-vending (and its

predecessor, the Inter-Branch-Card) renders the entire network of branches of a bank potentially accessible to the individual customer. To the extent that customers value this accessibility, the small scale entrant's ability to attract deposits will be correspondingly reduced.

These conjectures are based on our view of the direct impact of EFTS on individual depositors and the corresponding industry response. Such conjectures serve to motivate the need for a more complete model of the industry including empirically based estimates of the relevant parameters (elasticities) of the model.

Chapter 2 Selected Facts on the Canadian Banking Industry and a Selected Survey of Previous Work on the Economics of Banking

2.1 Canadian Banks

Extensive institutional information is already available on financial intermediaries in Canada. (Any undergraduate text such as Bond and Shearer is a source of such information.) The Canadian industry has five large banks with many branch offices (The Royal Bank of Canada, Canadian Imperial Bank of Commerce, Bank of Montreal, Bank of Nova Scotia, Toronto-Dominion Bank). Table 2.1 records the total Canadian financial assets held by all chartered banks and private non-bank financial intermediaries (including insurance companies, building societies, mortgage loan companies, trust companies, consumer loan and finance companies, credit unions, caisse populaires, Quebec Savings Banks, fraternal benefit societies, close-end funds, development companies, and mutual funds). The time series of the chartered banks' share of total assets is shown in Figure 2.1. From a high of 58% in 1945, this market share declines to a low of 36% in 1966 and re-emerges to a local high of 45% in 1980. Chartered banks including their entry are regulated by the Federal government through the Bank Act; near-banks may be either federally or provincially regulated. Recent changes in the Bank Act have permitted increased competition from foreign banks ('Schedule B' banks) although these banks are not allowed to compete through branching networks with the existing Canadian banks.

Some have argued that entry restrictions on banks (e.g., capital requirements) have resulted in a severe decrease in competition and that banking regulators are "captured" agents of the banking sector where the beneficiaries of banking regulation are the managers and shareholders of the bank. Even in this case, the monopsony power of banks as depository

Table 2.1

Market Share Data Calculated On

Canadian Assets Held by a Selected

Sample of Private Financial Intermediaries

	Chartered Banks		Private Banks*	
	Current dollars Millions	%	Current dollars Millions	%
1980	171 296	44.6	212 617	55.4
1979	147 285	44.3	184 949	55.7
1978	122 128	43.4	159 026	56.6
1977	102 819	43.0	136 572	57.1
1976	88 790	43.1	117 279	57.0
1975	77 169	43.3	100 954	56.7
1974	68 481	43.8	87 882	56.2
1973	56 455	42.0	78 104	58.0
1972	46 650	41.0	67 151	59.0
1971	39 958	40.0	59 958	60.0
1970	33 616	38.6	53 536	61.4
1969	31 000	38.2	50 124	61.8
1968	28 940	38.7	45 780	61.3
1967	25 199	36.8	43 267	63.2
1966	22 507	36.2	39 699	63.8
1965	21 196	36.6	36 776	63.4
1964	18 694	36.5	32 508	63.5
1963	17 858	38.4	28 712	61.7
1962	16 397	39.0	25 672	61.0
1961	15 644	40.2	23 260	60.0
1960	14 192	40.3	20 997	60.0
1959	13 442	41.4	19 029	58.6
1958	13 676	44.3	17 200	55.7
1957	12 274	44.0	15 650	56.0
1956	11 942	45.3	14 408	54.7
1955	11 575	47.2	12 956	52.8
1954	10 291	47.3	11 482	52.7

...2

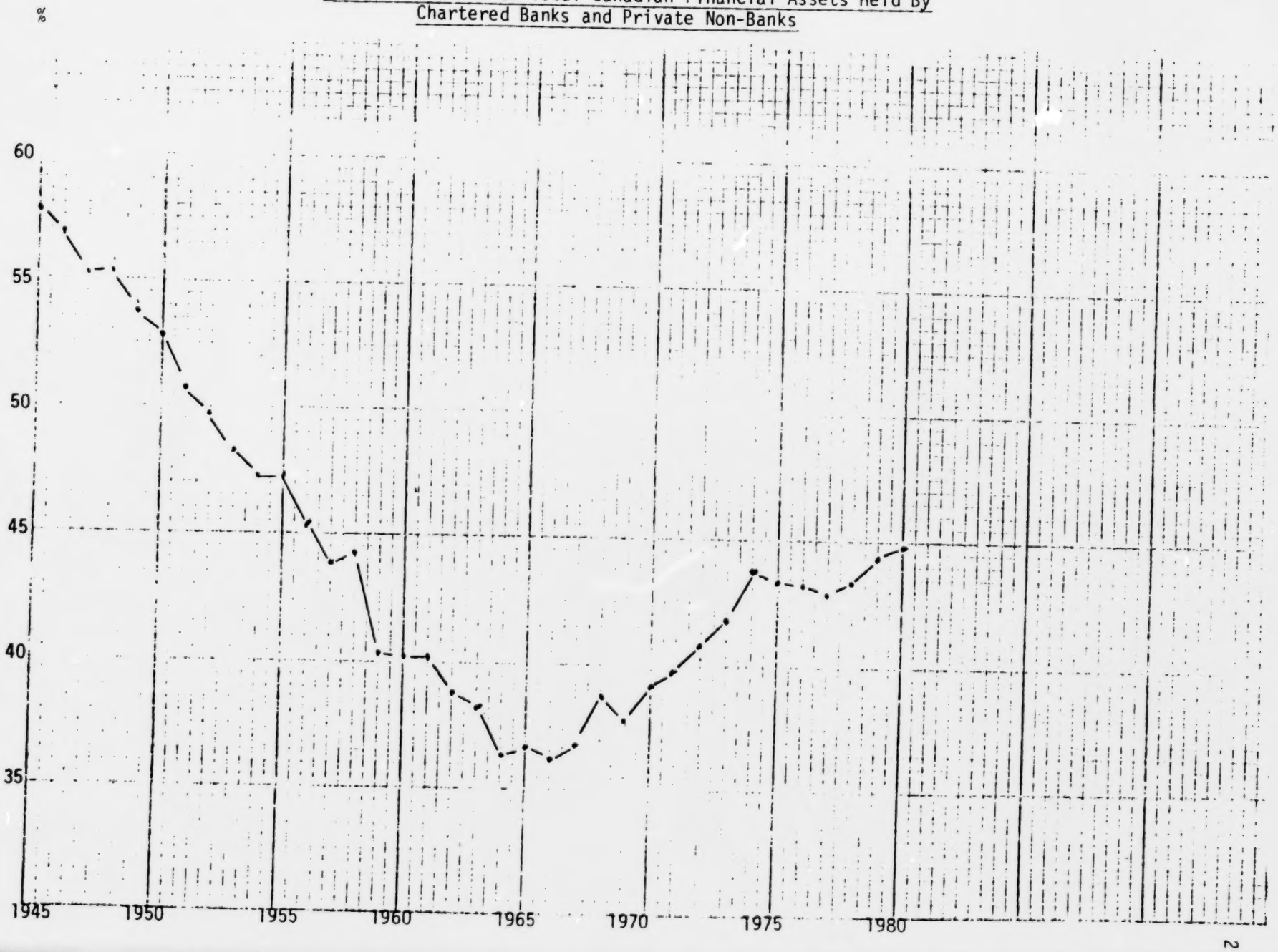
Table 2.1 (continued)

1953	9 592	48.2	10 290	51.8
1952	9 148	49.6	9 280	50.1
1951	8 589	50.7	8 354	49.3
1950	8 636	53.0	7 661	47.0
1949	7 853	53.6	6 796	46.4
1948	7 705	55.3	6 222	44.7
1947	7 067	55.1	5 764	44.9
1946	6 924	56.9	5 255	43.2
1945	6 517	57.6	4 807	42.5

*Private Non-Bank Financial Intermediaries includes insurance companies, building societies, mortgage loan companies, trust companies, consumer loan and finance companies, credit unions, caisse populaires, Quebec Savings Banks, fraternal benefit societies, close-end funds, development companies, and mutual funds.

Figure 2.1

Chartered Bank % of Total Canadian Financial Assets Held By Chartered Banks and Private Non-Banks



institutions has been eroded through competition from political jurisdictions. Provincial powers permit provincial authorities to regulate near-banks including conditions of entry. The competition for the quasi-rents from entry control between provincial and federal authorities may have eroded these rents. Nevertheless, our analysis in the previous chapter is predicated on some monopsony power in input price setting remaining with the chartered banks, a reasonable assumption if near-banks are not perfect substitutes for chartered banks. Further power may flow to the chartered banks as they are the only participants in the final stage of the Canadian cheque-clearing system.

2.2 A Selected Survey of Previous Work on the Economics of Banking

There is an extensive research literature in the economics of capital and financial markets and the nature of major institutions that trade in these markets. Some of this literature focuses on the problems of optimal portfolio management for financial intermediaries and the role of financial intermediaries as economic actors in the analysis of monetary policy. Most of this literature has little relevance for our purposes here. We may identify three objectives in surveying the extant economic literature on financial intermediation:

(1) the extent to which previous papers have modelled locational-branching decisions of banks in a useful manner;

(2) the extent to which previous models have specified the products and services produced by financial intermediaries that are relevant for an industrial organization study of the banking industry and the impact of EFTS in particular;

(3) the extent to which previous studies have modelled or estimated interesting features of electronic funds transfer systems.

The antitrust features of EFTS are developed in Chapter 3. Consequently, the literature reviewed by the authors directly bearing on EFTS is left to that chapter. Here, we will deal with the existing literature that impacts on the first two objectives stated above.

While there are many articles dealing with the banking firm itself, a recent paper that surveys the accumulative contributions of interest is by Ernst Baltensperger (1980). This paper falls into the second category of our list of objectives. While much of the material presented here is excessively detailed for the purpose of evaluation of EFTS and the nature of locational/branching decisions, it provides a useful framework and touchstone for the literature on banking.

Baltensperger begins by defining the functions of a bank as

(1) the consolidation and transformation of risks

(2) providing brokerage in credit markets.

Baltensperger identifies two specific types of risk - debtor risk or the risk of default in the asset portfolio and liquidity risk or the inability of the bank at any point of time to meet the demand for liquid funds by depositors. In particular, this paper has two major sections. In the first section, the paper analyzes the bank portfolio problem as an internal management problem for a fixed portfolio size. In the second section, the paper develops more complete models of banks emphasizing

(a) the possibility of monopolistic price setting by banks

(b) elements of risk aversion and their affect on bank behavior

(c) real resource models of bank behavior.

While elements of risk lie clearly at the heart of many decisions made by banks, we do not have to model explicitly risk for the purposes of identifying the equilibrium nature of the branching network. However, in the first

(a) the number of offices held by rival banks

(b) the initial distribution of accounts

(c) concentration ratios.

In particular, this paper is almost void of economic theory. Why age or history of a particular sector of the banking industry in the U.S. should matter is both unclear and not developed in the model.

A second paper (McCall and Peterson (1977)) is novel in two dimensions. First they had both an experimental and a control group. Secondly, they present a rather unusual interpretation of their results. In this paper, the authors had paired samples of rural banks. In one sample there was no entry (control group) while in the other there was entry (experimental group). The paper attempts to evaluate the impact of de novo commercial bank entry. To do so, the authors attempt to hold constant bank size, bank charter, the organizational structure of the industry, the competitive environment which includes the number of banks, the unit-branch mix, the presence of thrift institutions such as savings banks, and other non-banking structures, the market share of deposits held by the bank and the growth rate in the markets in question. Using thirty-nine measures of bank performance, the authors attempt to test whether there were differences in various performance measures of the banks in a period prior to entry, at entry and immediately after entry. Their results were that in non-metropolitan rural monopolistic or oligopolistic bank markets in restricted branching states, unit-bank entry had an impact on profitability and deposit interest-rate policies of existing banks. In particular, entry caused interest rates paid on time and savings deposits to increase significantly; at the same time the rate of return on capital experienced by existing banks in

section of this paper that deals with bank portfolio/internal management issues, Baitensperger presents a simple inventory model of liquidity management that is similar to the Baumol-Tobin model that underlies the model that we have developed in the first chapter. The paper then continues to deal with the problem of a bank in deciding on its reserves or alternatively its assets that can be allocated to earn interest and are therefore less liquid in the face of a stochastic outflow of deposits.

The paper continues to endogenize other elements of banking costs in a model of liability management. Here there are costs associated with issuing and maintaining deposit contracts. Such contracts require a real resource expense by the bank and therefore should be included in costs. These kinds of transactions costs are recognized explicitly in our model in the first chapter.

The remaining sections of this paper consider more detailed models of bank behavior and in particular deal with portfolio and reserve issues in the context of a richer specification of asset types. In general, this paper provides useful background material and specific elements of banking behavior including the specification of banking output are incorporated into our model.

There are several papers which purport to address the locational/branching decisions of banks. With one exception, the papers that we read were disappointing. They were disappointing in two respects - inadequate theory and inadequate measurement. Some papers (e.g. Ederington and Skogstad (1977)) deal with single-office banks. In the Ederington and Skogstad paper, their model establishes only that location matters by demonstrating that the price elasticity of demand for accounts in any single-office bank depends on

the presence of entry fell; finally in the sample market, banks did not engage in ruinous competition. The next effect is that with de novo entry, any branch bank market had less effect. None of these results are particularly surprising in light of conventional theory. The authors interpreted these findings to mean that existing banks via branching where possible had already engaged in significant competition through branching alone. Therefore, entry as a competitive response to above normal rates of return was a more significant phenomena in those geographical areas restricted by law to unit banks. In general, the authors conclude that the differential impact on bank behavior from entry may flow from the banking restrictions themselves.

Of the many other papers read, the one paper whose title held the most promise was a paper by Ali and Greenbaum (1977). In this paper, the authors develop an explicit spatial model of the provision of banking services assuming unit banking. In developing their model, they refer to Eaton-Lipsey's principle of "local clustering". Their model has a single FOB price (i.e., there is no price discrimination). The potential entrant picks the price and location. However, existing banks vary price only in response to entry elsewhere and there is sequential entry. They further assume that population is normally distributed over the real line and that consumers have identical tastes.

They use an exponential functional form for the demand function, common in this spatial literature. Costs to the consumer include, as usual, not only the costs of the banking services themselves but the transit cost to get to the local bank. (The cost to the consumer of banking is an important element in the model that we have developed in the previous chapter.)

For the banking firm, the costs include not only the cost of providing the service but the expected cost of obtaining a new bank charter for an entrant. This paper argues that the probability that an applicant holds of successfully obtaining the regulatory approval to enter the market is some function of the paid-in capital of the bank. The authors then define a minimum cost function with respect to the paid-in capital or size of the bank where the minimization occurs on a function of two elements - the sum of the paid-in capital itself and the expected cost of obtaining the charter. As we note above, this last item is perceived to be a function of the paid-in capital itself. An entry equilibrium occurs when the expected rate of return on these upfront costs just equals the anticipated discounted stream of profits that will accrue to the bank upon entry. Therefore, there is an open-entry phenomenon here based on the ability of firms to obtain a charter. The process by which charters themselves are granted is completely exogenous to the model.

There is extensive use of pseudo-empirics but in our opinion the results are obvious once they are matched up with the assumptions of the model. For example, the authors assume that when entry occurs at a previously occupied bank location, the existing bank(s) cannot changes its price. However, the new bank shares equally in the pre-established output. This means that with a cost structure that includes fixed elements, all of the banks at the particular location where entry occurred will face higher average costs. The consequence of this is that profits of both the existing banks and the entering bank fall. Therefore, with price invariant at a particular location even in the face of local entry, the model has no price game at all. The paper has a large amount of mechanical manipulations

with very little economic theory and once the assumptions are examined carefully, the results appear obvious. Therefore, while the title of the paper suggests the work would be significant to our research, the delivery of the material yielded only disappointment.

A recent study on limit pricing and branching is an unpublished Ph.D. dissertation by Dudley (1982). We comment here only on the empirical work contained in the last chapters. Dudley has two models, a deterministic and a stochastic model. Again, we will comment briefly only on the stochastic model.

This study focuses on de novo bank entry and uses banking data available for banks located in a rural Illinois communities. As Illinois is a unit-banking state, branching effects are not incorporated into this analysis. The study defines local markets and examines the differences between those markets that had entry and those markets that did not. While the data collected by the author covered a nineteen year period, the data are modified into a single observation for each market with all time effects removed. The statistical technique is a logit estimation of the probability of entry in each of the rural markets aggregated into a single observation through time. Explanatory variables include, amongst other things, the interest rate paid on savings deposits and the interest rate charged on consumer loans. Other variables attempt to correct for the size of the market and the annual growth rate of the market as well as dummy variables to account for the extent of competition in the market. As well, interaction effects between these exogenous variables were also included.

The author finds that consistent with the limit price model developed earlier in the thesis that in those markets where entry was effectively impeded by strategic behavior on the part of existing firms, higher interest rates were charged on loans and lower interest rates were paid on time

and savings deposits. While the empirical results are consistent with the limit price model hypothesized in the study, the author also recognizes that the results are consistent with a conventional competitive model of entry. Namely, those conventionally competitive markets where banks both charge higher interest rates on consumer loans and pay lower interest rates on time and savings deposits are markets where entry is to be expected to move the market from a position of short-run profits for firms to one where the long-run rate of return is conventionally competitive. Therefore, the difficulty with the empirical work is that the model cannot differentiate between sequential strategic behavior on the part of firms and conventional textbook competition.

The one final piece of empirical work that deserves special attention is the work on economies of scale in banking by George Benston (1972, 1976, 1982). There are two reasons for this special attention. First, Benston has made continued contributions to the empirical measurement of scale economies in banking. The estimation of such scale economies are an important ingredient into an understanding of the industrial organization of this industry. Such an understanding is an essential ingredient into any attempt to model the impact of technological change on the structure, conduct and performance of financial intermediation.

The second reason for paying special attention to the work by Benston is that the methodology is a standard methodology used to estimate the nature of production in banking. Benston estimates cross-sectional cost functions using accounting data and regressing costs against various measures of output and other exogenous variables. This methodology is subject to severe criticism and once we have presented briefly Benston's results we will evaluate the underlying estimation methodology. The same methodology

is used by D. Walker (1978, 1979) to estimate the nature of the production function for banks using various automated devices. These results are discussed in the next chapter but are subject to the same methodological criticism.

In their most recent paper Benston, Hanweck and Humphrey re-estimate cost functions for banks, both branch and unit state banks, using more recently developed functional forms. Their principle results are that average operating costs for both branch and unit state banks are U-shaped or upward sloping for three different measures of bank output. An optimum or minimum cost size of a bank office exists, is stable, and occurs for institutions with from $10,000,000 to $25,000,000 in deposits; the optimum size office is relatively small. Another conclusion is that branch banks appear to have a significantly greater potential to realize scale economies at the "plant" level. In contrast, larger unit state banks appear to face the significant scale of these economies. Benston et al. suggest that the average size of a branch banking office is small and around the size of a unit branch which experiences scale economies. But once the scale economy measure is altered to reflect actual branch bank behavior - expanding output by opening more offices as well as through output growth in each office - both branch and unit state banks experience significant operating cost scale diseconomies. This suggests that the "cost-efficiency" aspects of these two bank organizational forms are similar.

2.3 Some Brief Methodogical Comments on Cost Function Estimation with Cross-Sectional Data

Much of the work on the economic nature of banking institutions summarized above including the papers by Benston et al. and the empirical work on EFTS

by Walker summarized in the next chapter fits cost functions using cross-sectional accounting data obtained from individual firms. Several years ago, Friedman (1955) made several cogent points on the weakness of this methodology. First, cross-sectional data obtained from a sample of firms at a point of time may contain transitory errors. That is, at any given point of time, there is no guarantee that firms in the sample are in long-run equilibrium in terms of both the decistions made by individual firms and the competitive market forces that determine the nature of the long-run industry equilibrium. Furthermore, Friedman argues, errors made by firms in their profit-maximizing decisions will not be random, but rather the nature of contractual costs will leave the results from cross-sectional cost function estimation biased in favor of finding economies of scale where none exist. The reason is that firms may underwrite contracts with the suppliers of factors of production for periods of time in excess of any given market period. As a consequence, several firms may share the same cost structure yet be buffetted by stochastic forces which leave their outputs different. As a consequence, average costs will tend to be downward-sloping with firm size when, in fact, there may be no economies of scale at all in the production process.

Friedman highlights this regression fallacy by arguing that in long-run equilibrium, firms in a competitive industry even if they have traditional U-shaped average cost curves should be at one and only one point, namely the minimum point on the long-run average cost curve. Deviations in firm size from this efficient level of output are produced solely by random errors. Yet, the purpose of using cross-sectional samples is to capture firms in long-run equilibrium. However, if they were all truely at this

equilibrium output, there would be only one observation i.e., one output point for all firms and one average cost figure realized by all firms.

These difficulties are further compounded by measurement errors that occur in accounting data. Accountants and economists agree that arm's-length transactions conducted in a market produce reliable cost estimates. The difficulty arises from the accountants' measure of depreciation on fixed assets and the accountants' measure of intangibles such as goodwill. Fortunately, in the banking industry, the depreciation on fixed cost does not constitute a major item in the expense portion of the income statement. Nevertheless, items recognized by economists as legitimate costs may not appear in traditional expense sections of the income statement.

These comments advanced by Friedman do not necessarily hold in a world of imperfect competition with product differentiation i.e., non-price competition. The reason is straightforward. Even though firms use the same underlying production technology, the presence of non-price competition and other informational costs mean that firms, having established their niche in the product marketplace, will incur varying costs associated with the specific product they supply to the market. Thus, banks or other financial intermediaries that provide more service (e.g., more branches) or more information to their customers (charging a higher price to cover the costs of these services) will incur higher costs associated with the provision of these services even though the costs to all banks of processing items such as cheques, deposits and consumer loans may be relatively undifferentiated. As a consequence, banking and other financial intermediaries in equilibrium may have different sizes of output, charge different prices, incur different costs even though the rate of return to all financial intermediaries may be equal.

Chapter 3 Economic Features of EFT: Competition and Cooperation

3.1 Introduction

While much of the extensive trade literature on the impact of EFT on the behavior of financial intermediaries in general and banks in particular provides useful data on the nature of EFT equipment, most of this information lacks a suitable framework for economic analysis. In earlier chapters, we attempt to provide such a structure. Retail banking can be best understood in terms of a spatial monopolistically competitive model. Banks and their competitors locate with a view of maximizing their profits. Profits are maximized in equilibrium when banks locate so as to service the needs of their consumers (where service incorporates the time costs of consumers for access to this banking system) in a fashion that may simultaneously preclude competitors from entering so as to dissipate all economic rent. Rent dissipation may be prevented because of location specific fixed costs. On the one hand, EFT represents a change that has the potential for dramatic impact on the structure of financial intermediation because it removes the monpolistic element of banks and their competitors due to the locational features of behavior. This pro-competitive effect must be balanced against an anti-competitive effect that arises because EFT appears to be characterized by significant economies of scale. Previous chapters have emphasized that EFT should be viewed as a form of technological change. However, it is the presence of significant economies of scale that differentiate this form of technological change from technological change which represents a downward shift of perfectly elastic supply functions in a world of competition undistinguished by locational monopolistic features. In a competitve environment, the profit incentive manages to induce firms to introduce cost-saving advantages in an effort

to collect short-run profits. Incentive problems may arise when successful innovations can be immitated at a zero cost by other firms in the industry. This appears not to be the case for the banking industry for the main thrust for the electronic technological innovation comes from outside the banking industry, namely the computer industry itself.

The central issue on the impact of economies of scale on the organization of financial intermediation concerns the magnitude of the scale economies. Are these economies sufficiently large that a monopolistically competitive industry can be replaced by a single monopoly? Or in the presence of fixed costs that are large enough to be relevant but not so large as to permit a single firm to emerge, what coalitions will form to realize the efficiency gains from size? In the coalition formation that follows from the introduction of EFT, will there be sufficient coalitions to yield a degree of competition that guarantees that the benefits from the introduction of cost saving devices associated with EFT are passed along to the consumer? These questions require a successful estimation of the magnitude of the economies of scale from the introduction of various electronic funds transfer devices. Some would argue that in contrast to the United States where regulation at the state level frequently forbids branch banking, Canada has by some standards an industry already characterized by significant degrees of concentration. In order to understand fully the economic issues associated with an evaluation of the imapct of EFT, it is necessary to marry an understanding of the forces that have shaped the current structure of the branching networks of the major banks in Canada with the economic features of an EFT technology. We discuss the nature of the banking production function including the associated cost of branching in the

previous chpater. Here, our focus will be on the nature of the economies of scale inherent in EFT and an evaluation of the economic forces that will determine the structural changes realized by the Canadian industry. In the final chapter, we will synthesize our understanding of the forces that determine the economic consequences of EFT with those that determine the nature of the existing bank structure in Canada. We will evaluate these forces in light of the theoretical model of the first chapter and comment on the work to be done in the future.

3.2 Economies of Scale and EFT

How large are the economies of scale that flow from the various technological innovations that are broadly labelled as EFT? The evidence at hand is drawn from two studies by David Walker (1978, 1979). Both studies comment on an empirical estimation of cost functions of electronic funds transfer devices from a cross-sectional sample of 534 U.S. commercial banks sampled in 1974 that had installed various EFT equipment. In particular, the devices that are Walker's focus in both papers are cash dispensers and automated tellers; Walker recognizes that other automated devices include point-of-sale terminals (these constitute retail elements in the system) and automated clearing houses, regional switching centers and other connecting elements (these constitute wholesale elements in the system). When analyzing the cross-section cost function estimates of banks in the previous chapter, we comment on the limitations of this estimation technique using conventional accounting data. These caveats apply with equal force to the cross-sectional cost function estimates performed by Walker.

In particular, Walker estimates cost functions for cash dispensers and automated tellers. Regressions using ordinary least-squares estimation techniques reported by Walker seek to explain the corresponding costs for these two items as functions of a measure of electronic funds transfer output, non-electronic funds transfer costs and exogenous factors. It is traditional in cross-sectional cost function estimation to ignore factor prices by invoking the assumption that all firms compete in competitive factor markets for inputs and therefore face identical factor prices. Consequently, factor prices are controlled for and any observed variation in cost cannot be due to some variation in factor prices. The inclusion of cost for items other than the electronic funds transfer devices on the right hand side of a cost function regression is an unusual specification. It assumes a separability between the electronic and non-electronic cost items that may not in fact hold. For the purpose of cost minimization, it is traditional to assume that outputs are held fixed and therefore are exogenous to this partial decision process. This is not the case for other inputs. Rather, the traditional approach would explain all cost items as functions of exogenous variables (including factor prices) and measures of output. It is not possible to determine what biases may be introduced into the empirical results from this non-conventional specification.

Walker includes eight exogenous factors:

1. whether or not the bank is located in an urban area
2. the geographical region in which the bank is located
3. whether or not the bank has more than two offices
4. whether or not the bank is a member of a bank holding company

5. whether the institution is a national bank, state-member bank or state non-member bank
6. the size of the bank as measured by its total assets
7. whether or not the bank shares its retail EFTS machines with another financial institution
8. whether or not the bank operates its retail EFTS machines in an "on-line" environment.

These factors were included to take account of differences in banking legal environments, locational differences and other structural influences on EFTS operations.

The two activities that form the focus of Walker's estimation are cash dispensers and automated tellers. Successfully estimating the correct corresponding cost functions for each of these two items should permit us to determine whether there are significant economies of scale associated with size. Walker uses a logarithmic specification of the cost function which, if nothing else, permits a simple interpretation of the coefficients. Walker's estimates indicate that there are significant economies of scale for both cash dispensers and automated tellers. Scale is measured by total transactions volume where the transactions activities include cash withdrawals from chequing accounts, savings accounts, and advance on credit cards; deposits to chequing and savings accounts; account transfers between chequing and savings accounts and from credit card accounts to chequing accounts; and payments on mortgages, bank loans and merchant accounts. For a 1% increase in this aggregated total transactions volume cash dispenser costs increase by only .12% and automated teller costs increase by only .50%. It is not possible to tell from this specification where these economies of scale cease to be significant.

By taking the range of output for banks in the sample, Walker is able to calculate average maximum and minimum costs per transaction. There are banks in this sample who can realize average cost of 10¢ per item for cash dispensers and 11¢ per item for automated tellers. This is contrasted with an average cost of a typical transaction through a bank branch at 27¢. Both of these cost items reflect variable costs and do not include fixed or overhead costs. (Walker reports that the estimate of an average cost for a non-automated bank comes from a survey of 14 large banks by the American Banking Association.) These estimates raise two issues that will be dealt with in the remainder of this chapter. First, if coalitions form that exclude some financial intermediaries that operate branches, take deposits and issue loans, will these intermediaries be excluded from the market? Second, if these "common-carrier" characteristics of an automated EFT system are significantly large, what pricing principles and/or regulatory devices can or should be imposed by public authorities? We now turn our attention to these issues.

3.3 Antitrust Considerations

The theoretical model specified and analyzed in Chapter One identifies several opposing forces that flow from the further introduction of automated devices in banking. Banks have improved and centralized information on the credit worthiness of consumers; cash dispensers, automated tellers and point-of-sale terminals are more densely located. Such changes reduce both the role for conventional bank branches and any monopoly elements in the current branch bank structure as well as having implications for entry. (We have previously argued that any monopoly elements in the branch structure come from location-specific costs associated with branches. In the presence of such costs, banks can credibly locate branches in the retail market space

to preempt entry, leaving an equilibrium where each branch may realize positive economic profits.) Informational costs have traditionally supported market segmentation and specialization on the part of banks and near-banks. For example, trust companies and insurance companies traditionally played large roles in mortgage markets; banks traditionally played large roles in consumer demand loan markets. An improved and automated information technology which lowers the costs for centralized information networks should eliminate the informational advantages that facilitate such market segmentation. Therefore, all financial institutions should become more alike and compete more directly with each other. In general, improved cash management for consumers should permit lower transactions cash inventories and heighten competition on the part of financial intermediaries for savings and demand deposits by consumers. These factors all weigh in favor of an increased competitive environment in the market for financial intermediation. In this chapter, we have identified scale economies in EFT that may loom large relative to the size of the market place. Scale economies produce restricted competition. Competition is restricted because firms that are large can afford to charge lower prices because they realize lower costs than firms that are small. The crucial question is whether or not scale economies weigh in favor of firms that are large relative to the total demand for the product. This factor weighs in favor of a decreased competitive environment.

The crucial question is on net which forces dominate - forces that enhance or restrict competition? Even if the degree of competition were to remain unchanged, EFT would alter the nature of an efficient banking

institution: Branches as they exist will decrease in number; the density of automated equipment will increase, more than replacing the existing branch network; if transactions costs are lower with an increased density of machines to access the branch network, consumers will benefit; the minimum scale of a banking institution for survival will increase; surviving banking and near-banking institutions will be more similar and compete more directly for consumer savings deposits. Figure 1.1 (in Chapter 1) plots the number of branches per capita for chartered banks from 1945 to 1981 for Canada, British Columbia, Ontario and Quebec. While the averages are different, all reveal one pattern. Branches per capita grow or remain constant for a long period but have shown recent signs of a decline, for Quebec from 1966, for the rest of the provinces shown and for the country as a whole from 1975. Are these the early effects of EFTS on the branching network?

In the remainder of this section we will focus on the antitrust issues associated with the scale economies that accrue to surviving and dominant institutions from automated equipment. In particular, we will develop two simple analyses to illustrate the trade-offs that will determine the final equilibrium nature of the market. These trade-offs will illustrate the type of considerations and evaluations that antitrust authorities should pursue to evaluate properly the outcome of the introduction of electronic systems. In the first illustration, we will ask what orders of price elasticities of demand for banking services will have to exist in order to support a banking system where some firms have access to lower costs through automated equipment while other firms remain excluded from the electronic networks that will form to realize cost advantages. In particular, when will the market support two types of firms? Alternatively, when

will the market permit those firms that are part of a coalition that realizes lower costs through EFTS to eliminate completely other competing firms excluded from the new technology? The second illustration will focus on the welfare and efficiency trade-offs that will occur as a consequence of the introduction of cost-saving technologies with economies of scale. In general, if technological innovation with economies of scale permits surviving institutions to exercise monopoly power and charge prices higher than marginal costs, there will be the usual deadweight welfare loss associated with monopoly pricing. However, the technological advance will save resources for those units of output actually produced and sold through the marketplace. Therefore, if we are willing to add-up dollars of benefits and costs independent of the beneficiaries and losers involved, we can trade-off welfare gains (the savings of resources) against welfare losses (the deadweight loss from monopoly pricing).

Illustration 1

This illustration follows the development in Dixit and Stern (1982). Ignore the explicit spatial features of branch banking for the moment and assume that the market for financial intermediation is an undifferential oligopoly. Define the inverse demand for the simple service of this industry (e.g., accounts) as;

$$P = \Phi(\Sigma X_i)$$

where P is price and X_i is the output of firm i (e.g., bank, near-bank). In this oligopoly, assume that conjectures of the participants can be represented as constant conjectural variation of the type $(dX_k/X_k)/(dX_i/X_i) = \alpha$.

In this case,

$$d(PX_i)/dX_i = P[1 - (\alpha+(1-\alpha)S_i)/\varepsilon]$$

where $S_i \equiv X_i/\Sigma X_i$, the market share of the i^{th} participant;

$\varepsilon \equiv (\phi' X_i)/P$, the market demand elasticity.

Individual profit maximization means that $d(PX_i)/dX_i = c_i$ where c_i is the marginal cost to the i^{th} participant.

Therefore, combining these elements yields:

$$(P-c_i)/P = [\alpha + (1-\alpha)S_i]/\varepsilon \qquad (*)$$

Imagine that as a consequence of the availability of EFTS with significant economies of scale that a long-run equilibrium exists where some firms (e.g., banks) have access to the automated system while other firms (e.g., near-banks) do not. Banks could charge a fee to the near-banks for accessing the EFTS just equal to the opportunity cost for the near-banks.

Walker (1979) reports that the minimum average variable cost for the largest firm in his sample for cash dispensers is 10¢ per item and for automated tellers is 11¢ per item as compared to a non-EFT operation where the average variable cost is 27¢ per item. Approximate marginal cost by an average variable cost of 10¢ for EFTS and 27¢ for non-EFTS.

Information on the potential structural impact of EFTS on the market for financial intermediaries under our assumptions is generated by using the price/market share/elasticiety/conjectural variation relationship (denoted above by (*)) to calculate the market shares of banks and near-banks, using the cost data reported in Walker, that correspond to alternative estimates of α and ε . Notice that the market structure underlying (*) is <u>non-strategic</u> in the sense that entry is not sequential nor do firms solve their maximization problems knowing their ability to "knock-out" rivals.

With undifferentiated oligopolies, all firms face the same equilibrium price so that a reformulation of (*) together with the knowledge that $S_2 = 1 - S_1$, where $S_{1(2)}$ represents market share for firms with (without) access to EFTS, yields:

$$S_1 = [\varepsilon(c_2 - c_1) - (\alpha c_2 - c_1)]/[(1-\alpha)(c_1 + c_2)] .$$

If we assume Nash conjectures (i.e., $\alpha=0$), the following market shares corresponding to given demand elasticities may be calculated:

S_1	ε
.75	1.05
.85	1.25
.96	1.50
1.00	1.58

The result is that price elasticities do not have to be very large before the banks with EFT completely eliminate their less efficient near-bank rivals without EFT. Estimates indicate that the market share of banks in recent times is approximately .45. (See Table 2.1 in Chapter 2 for historical market share data in Canada. Modest conjectures on demand elasticities could permit banks in our illustration to double with ease their collective market share.

Nash conjectures are most favorable to a high market share for near-banks. In fact, it is straightforward to calculate that $\partial S_1/\partial\alpha = (\varepsilon-1)/(1-\alpha)^2 > 0$ (as $\varepsilon > 1$). The economic reason for this result is equally straightforward: The more elastic are any firms conjectures about its rivals' responses (greater α), the more elastic is the demand curve faced by the individual firm. Firms with EFTS and lower costs will mark-up their costs less, (have less monopoly power), the smaller their conjectured price elasticity. However, the lower the final equilibrium price, the closer the high cost,

non-EFT firms will be to elimination from the market. By setting $S_1=1$, the critical ε that eliminates near-bank competition may be calculated as $\varepsilon^* = c_2/(c_2-c_1) - c_1/(c_2-c_1)\cdot\alpha$. From this relationship, it is obvious that $\partial\varepsilon^*/\partial\alpha = -c_1/(c_2-c_1) < 0$ (as $c_2 > c_1$).

One caveat on the underlying model is that with the clear domination of one group of firms (those with EFT), the others (those without EFT) may become price takers so that the appropriate model may be one of a dominant firm(s) (banks) with a competitive fringe (near-banks). In this setting, the competitive fringe are passive price takers, setting price equal to their marginal cost. Obviously, if such a competitive group cannot cover costs in the long-run, they fail to survive and the dominant firm is left as a monopoly. Are the fixed costs and scale economies from EFTS sufficiently large to yield these results? This question cannot be answered in the absence of more detailed estimation of the relevant parameters.

Illustration 2

This illustration supposes that a pre-EFTS equilibrium involves zero industry profits. This condition is somewhat at odds with the theoretical model of Chapter 1 where the initial branching equilibrium is characterized by positive profits (flowing from the sunk locational costs). However, the purpose of our theoretical model is to provide a prototype model of a branch decision structure for a bank that highlights those service elements of the bank's output and structure relevant for analyzing the impact of EFTS in a setting where the entry pattern is critical. There are several factors even within this setting that could yield zero profits.

Banks are multiproduct firms. The model of Chapter 1 focuses on price (interest rate) and locational competition. Even if monopoly elements in location and price leave positive profits for banks, other non-price features of bank competition (e.g., advertising, give-aways for new customers, travellers' cheques) may dissipate these rents. If the location-specific costs of branching leave price above the marginal cost of producing the bank's output and the marginal cost of other non-price competition is constant, banks may compete in non-price variables until the profits are dissipated. (This general point was made by Stigler (1968).)

A second possibility for rent dissipation exists because banks knowing that early entry with location-specific sunk costs has the potential to pre-empt subsequent entry by competitors will establish a branch once the discounted profit stream is non-negative. Consider an expanding suburban area. Establishing a branch in a critical location may produce monopoly elements that ultimately yield positive profits in future years once the surrounding market area is completely established. However, the desire to capture these future rents will encourage banks to establish branches to pre-empt rivals once the early losses (from branching before the density of customers is sufficient to yield break-even profits) are just offset (in present-value terms) by the future profits. From a cartel or monopoly profit of view where such entry externalities would be internalized, entry occurs "too early". (Similar points on (i) the timing of innovation and (ii) early entry were made by (respectively) Barzel (1968) and Eaton-Lipsey (1979).)

In a post-EFTS equilibrium in this illustration, suppose that firms with EFTS (banks) "knock-out" the non-EFTS competition (near-banks) and, as a consequence, act as a cartel and possess market power because of the significant scale economies of EFTS. Again, we do not model spatial features or other forms of non-price competition.

In this admittedly naive setting assuming a competitive pre-EFTS world, the welfare trade-offs are simple: efficient pricing in a pre-EFTS world versus inefficient pricing but an improved technology and therefore resource saving in a post-EFTS world.

The relevant formula may be borrowed directly from Williamson (1968). Net welfare from EFTS is positive in this simple setting if

$$\Delta AC/AC - .5\varepsilon\ (\Delta P/P)^2 > 0$$

where AC is average cost, P is price, ε is demand elasticity and Δ indicates change.

Or by substituting the estimates of relative costs and assuming a price elasticity of 2 (sufficient to "knock-out" non-EFT rivals according to our previous illustration), the critical price increase that may be tolerated before net welfare becomes negative is 79%, a substantial price increase.

While a similar welfare evaluation may be conducted using the model of Chapter 1 the results need not be the same. First, neither this calculation here nor the one in the previous illustration include the non-price effects of the technological change. If machines more than replace bank branches, it is difficult to see how the cost of access to the system could not be reduced for the consumer. Such changes, therefore, add to the potential benefits from the change as well as the resources saved by the market participants (i.e., $\Delta AC/AC$). The elimination of some competition together with the size constraint for efficient entry may reduce conjectured price elasticities sufficiently for the efficient survivors that prices may ultimately rise but the question is: how great is the price increase? In the absence of a fully-developed welfare analogue to our earlier theoretical model, this question remains unanswered.

3.4 Regulatory Principles

The previous sections in this chapter and elsewhere have identified the potential economic forces at work to alter the structure of financial institutions from the introduction of various automated devices into banking. Some of these effects are pro-competitive while others are anti-competitive. Do the appropriate incentives exist for existing private institutions to introduce the pro-competitive effects in a manner that avoids anti-competitive consequences? Institutions are, of course, motivated by self-interest. The answer to the question, therefore, turns on the discipline that emerges from the marketplace. Where possible in our previous analysis, we have tried to identify the precise nature of the relevant market incentives.

Where the market fails in some sense, what precisely, if anything, should the public sector do about it? We have identified several sources of potential market failure: location-specific sunk costs of branches that may facilitate sequential market preemption by existing firms yielding some monopolistic market power; in turn, such monopolistic power may yield a reluctance on the part of banks to introduce automated devices (networks) that are inter-bank in nature as such devices may erode existing market power; significant economies of scale in EFTS, documented for cash dispensers and automated tellers, that may yield market power to banking coalitions conditional on the magnitude of these scale effects, and the cost of coalition formation - if sufficient EFTS coalitions exist, competition among networks may preserve sufficient levels of competition that the majority of the benefits from such technological change accrue to consumers.

In this section, we turn to a brief discussion of the regulatory issues that may be important for policy makers should the introduction of EFTS produce a cartel in the provision of these automated cleaning-house services.

There is a vast theoretical and empirical literature in economics coupled with extensive legal documentation on the nature and experience with regulatory intervention. The public choice literature in economics has tackled the challenge of developing positive theories of instrument choice to explain cross-national differences in regulatory institutional structure - why do governments choose to produce the output or service rather than contract on its provision, i.e., why does Canada relative to the U.S. use Crown corporations rather than rate-of-return regulation of privately-owned corporation or taxes and subsidies. This literature while ultimately important for any policy choice is sufficiently removed from the current issues that in our brief discussion of pricing issues we ignore the specific institutional structure that might be in place.

Suppose that the introduction of a massive automated system leaves only one viable coalition that has the power to eliminate non-member rivals. What is the appropriate public policy response? Some economists (e.g., Demsetz 1968) would argue that little should be done. The public sector should recognize that the magnitude of the scale economies will leave only one (perhaps a cartel) producer of the services but that scale economies do not preclude competition for the right to be the sole producer of the product. Therefore, an auction by the public sector for the property right to be the exclusive agent operating the centralized EFTS where the winner of the auction is that firm with the lowest price would be the bidder

offering to set a price equal to long-run average costs. While average cost producers are not marginal cost producers, Demsetz conjectures that such an auction would be efficient and first-best if transactions costs are included in the cost figures.

An alternative to this arrangement would be a government-owned or government-regulated utility given the exclusive right to develop and operate the centralized EFTS in perpetuity. Of course, it is well-known that in the presence of significant scale economies, an efficient utility setting price equal to short-run marginal costs (with the correct investment rule i.e., expand capacity until the cost at the margin just equals the savings in variable resources) would earn negative profits. Such losses would have to be covered by subsidies or a departure from marginal cost pricing.

Other political considerations may dictate the final form of the institutional structure. Or the grounds that politicians are more facile than academic economists at determining the optimal political strategies that maximize the chances for survival, we offer no economic analysis of these choices.

Chapter 4 A Summary of the Principal Results

We distinguish two general forms of electronic funds transfer and two forms of access by individuals to the services of commercial banks. These distinctions are essential to an understanding of the economic consequences of introducing EFTS into the banking sector. First, electronic funds transfer can take the form of intra-bank transfers, cash-vending for example, or the form of inter-bank transfers, point-of-sale transfer for example. Second, individuals can access the services of a commercial bank either directly, using a bank card for example, or indirectly by withdrawing cash for example. Intra-bank EFT tends to reduce the individual consumer's cost of indirect access while inter-bank EFT tends to reduce the cost of and to increase the scope for direct access. In general, these forms of EFT operate through different channels; in particular, the introduction of inter-bank EFT is necessarily a collective decision while intra-bank EFT is an individual decision. Therefore, the implications of these two forms of EFT and the prospect for their introduction are conceptually different.

Our analysis indicates that intra-bank EFT tends to increase both price competition within the industry and the scale necessary for entry to the industry. As a consequence of the introduction of this type of EFT, we expect a reduction in the number of branches within the industry - fewer branches are required to deter entry and the profits of individual branches decrease as intra-bank EFT is introduced. This expectation is consistent with the recent decline of branches per capita in Canada. (This decline began approximately in 1975.) The implications for overall profitability of the industry and for the welfare of consumers are ambiguous. Informed conjecture on these issues requires a more elaborate and carefully callibrated and measured model.

Inter-bank transfer of funds raises some very intriguing and interesting problems. Our view of the source of rent within commercial banking is that, for the most part, it arises from spatial monopoly and monopsony power. If, as seems the case, the introduction of inter-bank transfer were to increase significantly the direct access of bank customers to bank services, then any spatial market power would be substantially eroded and as a consequence, so would the potential for significant rent within the commercial banking industry.

However, if such an inter-bank electronic funds transfer system were characterized by significant economies of scale (as currently conjectured and measured), then this potential rent in large measure would be transferred to the owners of the inter-bank transfer network. Of course, this current existing rent would be increased to the extent that the introduction of an inter-bank EFT system realizes a dramatic reduction in unit cost. The obvious candidates to be the innovators and therefore owners of such a network are the commercial banks themselves. If these banks anticipate that they could realize this potential rent from the introduction of such a network, we conjecture that they have a significant incentive to introduce the network at some point. Further, if the technology is as efficient as many observers claim, the introduction of the technology even with a high degree of concentration in the ownership of such a system would seem to leave consumers better off as well. (Substantiation of these conjectures requires, of course, a more detailed economic model.) Further, if the ownership coalition for such an extensive network were to exclude near-banks and near-banks collectively were not sufficiently large to undertake the investment required to duplicate an efficient network (or were explicitly precluded from doing so because of regulation) then such near-banks would cease to be

effective competitors for the working deposits of individual consumers.

In assessing the prospects for the introduction of this innovation, our view is that the crucial question is: To what extent would public authorities allow the owners of an extensive electronic funds transfer system to extract the rent that would acrue from such ownership rights? We may restate this question from the point of view of the innovators to highlight their incentives for introducting such a system. In that case the crucial question can be restated as: How do the commercial banks perceive the regulatory environment that they would face after the introduction of an extensive electronic funds transfer system? If the perception of the commercial banks is that little or no rent would be realized from the introduction of the network because of the antitrust implications of such a network and the anticipated action by public authorities, then we conjecture that the banks themselves will not innovate because of their perception of the nature of a long-run equilibrium. Furthermore, others will not likely introduce such a system since the effective operation of an inter-bank network requires the cooperation of the commercial banks themselves. This raises the possibility that while the individual actors in the market for financial intermediation may have a sufficient incentive to introduce automated devices that impact solely on their own operations, the more extensive networks that require collective action may never appear as the individual institutions do not have the incentive to introduce a technology that appears to be superior (less costly) for the transfer of funds.

From this perspective, we can identify three important research tasks to be undertaken. The first is the development of a more detailed model of this industry along the lines of the model sketched in Chapter One of

this report. The second task is a careful estimation of the parameters of that model in order to specify its comparative static properties. (Where warring forces exists in the model as a consequence of a parametric change, which forces dominate?) Although the third task appears somewhat more academic at first blush, its execution is critically central to our approach. The central and critical assumption to our model is that banks manipulate sunk costs strategically. That is, our model employs the new industrial organization paradigm discussed at length in an appendix to this report. Do commercial banks behave this way? We propose to investigate further this question by examining the establishment of new branches by commercial banks in growing markets. In such markets, this paradigm has an unequivocal prediction: Branches will be established by commercial banks in growing markets at points in time where these branches do not cover their costs. Such early entry accompanied by the committment of sunk location-specific costs effectively substitutes for a property right. Such an empirical exercise is feasible due to the existence of detailed census-tract data. Detailed micro data should permit us to identify both growing and stable or mature markets. Such identification facilitates an experimental and control group. Whatever the form of a reduced-form equation that would flow from a fully specified model of the type sketched here, the critical empirical question would be whether the branching equation for growing markets was statistically different from the branching equation for the stable markets. The successful estimation of such equations should permit us to identify both the size and existence of economic rents.

Appendix:

STRATEGIC BEHAVIOR IN OLIGOPOLIES:

A LITERATURE SURVEY

Our objective in this appendix is to survey and synthesize the existing literature on strategic behavior on oligopolies. Section 1 presents an overview of the relevant economic literature. Section 2 examines two simple models of strategic behavior in a duopoly setting. The purpose here is to convey an understanding of the essential central ideas in the economic literature on strategic behavior. Chapter 1 of this report draws on this literature to develop a model of the commercial banking industry.

I. An Overview of the Theoretical Literature

Strategic free-entry is best understood through two basic characteristics. The first fundamental distinction in the paradigm of strategic free-entry equilibrium is between variables whose values are irreversibly determined prior to marketing of the produced good and those variables which cannot be irreversibly determined prior to marketing (Dixit 1982). Irreversibility is the key. Some examples will serve to illustrate this property. Product development costs are costs that must be incurred prior to the production and marketing of a product. Once the resources have been committed and used for this end, they cannot be reversed. Similarly, the physical plant of a manufacturing firm, created prior to production, is irreversible as the plant cannot usually be physically reconverted back into those basic resources (inputs) used to product it in the first place. In contrast, product prices do not share this property of irreversibility. Although product prices may

be provisionally determined and announced prior to production and marketing, they can be changed when the goods are produced. With the goods in hand, firms will set prices to maximize profits conditional on the nature of the market at the time of sale.

A second fundamental idea in the strategic literature is the notion of specificity of variables. Again this idea is best illustrated by an example. Consider the expenditure of a firm on any item used in the production and sale of a product. The specificity to that firm of any particular variable can be measured by the maximum value of that variable in some other use. In particular, the ratio of the original expenditure to the value of that item in its alternative use is an index of specificity. (This index ranges from one or no specificity, to infinity or complete specificity.)

Specificity and irreversibility are closely related concepts. Indeed, specificity is necessary for irreversibility: Completely non-specific items can be converted to their best alternative use at no cost penalty, while completely specific items have no value in any alternative use and the creation or production of such items is irreversible.

A useful separation is the partition of variables for a firm into two sets, those chosen prior to marketing which are irreversible (or alternatively specific), and those which are not irreversible. We will refer to the first set of variables as precommitted variables (or committments) and to the second set as non-committed variables. This partition of variables imposes some strong restrictions on the form of any economic game involving competing firms. Consider the game played by competing firms when their goods are marketed, labelled the market game. When this market game is played, the values of precommitted variables are fixed and unalterable. The range of choice open to firms in this game is then restricted by previous decisions with respect to precommitted variables. This implies that any market game

is preceded by still another game in the precommitted variables, which we label the committment game. Costs are incurred in the committment game, but revenues are generated only in the subsequent market game. Thus, in the committment game, the rational firm is forced to consider the relationship between committments created today and net revenues which accrue to the firm in tomorrow's market game. The first innovation from the strategic models is the recognition of the sequential nature of such games. The key reference is Prescott-Visscher (1977).

The second innovation from these strategic models relates to the nature of the dependence of net revenues in the market game on the values of precommitted variables. In order to make rational decisions in the committment game, the firm requires a theory or a conjecture about this dependence. The salient assumption about such conjectures is that they be rational or correct. The logic is simple but compelling: If we do not constrain our theorizing by imposing the restriction that in making committments, any firm's conjectures on the market game be correct, then virtually any conceivable outcome of the two interconnected games is possible. Any resulting theories would be completely uninformative. This view of the problem is explicit in Eaton-Lipsey (1979) and implicit in Prescott-Visscher (1977).

The third innovation in this literature is the realization that firms which have the opportunity to choose committments have both the ability and the incentive to manipulate these committments to create asymmetric market equilibria in their favor. (The most extreme form of this asymmetry is preemption, the focus of much of this literature.) The key articles are Eaton-Lipsey (1979), Salop (1979) and Schmalensee (1978), although the basic insight is apparent in Archibald-Rosenbluth (1975) and Eaton (1976).

The emphasis on preemption is unfortunate since the resulting impression (mistaken we believe) is that this literature is concerned exclusively with monopoly rents and their protection. The fundamental messages in the strategic literature have much wider application. The sequential view of equilibrium driven by the dichotomy of choices into committed and non-committed variables promises a general theory of market structure in which asymmetry is the rule and not the exception. Monopoly preemption is then a simple and particularly illuminating special case. From this point of view, Dixit (1980) and Eaton-Lipsey (1980) are worth reading. The general theory of market structure implicit in this literature is only now emerging. We believe that the Eaton-Mathewson project on EFT and banks will be an important step in this direction.

II. Some (Non-banking) Illustrations

The force of these basic ideas is most efficiently illustrated through some examples. Here, we present two examples, the first has research and development expenditures as the precommitted variable (the example is inspired by Brander-Spencer (1982)), the second has advertising as the precommitted variable. Both examples consider a duopoly. To focus on the causal mechanics driving the asymmetric results in the market game, we ignore the possibility of entry preemption at the committment stage. (The extension of our illustrations to entry preemption is straightforward.)

In the simplest case, research and development reduces costs of production. Let $C(X,R)$ be the minimum cost of producing output X, given that R has been devoted to cost reducing research and development. By

the definition of R, $\partial C(X,R)/\partial R < 0$ and we assume that $\partial^2 C(X,R)/\partial X \partial R < 0$ i.e., increasing R reduces both the total and marginal costs of production. In addition, we assume that $\partial^2 C(X,R)/\partial R^2 < 0$, i.e., these cost-reducing effects are diminishing at the margins.

Now consider an undifferentiated duopoly problem in which (i) firm 1 chooses R_1 first, (ii) firm 2 subsequently chooses R_2 and (iii) the two firms then play a market game in which their decision variables are their respective outputs, X_1 and X_2.

Observe that in the market game, stage (iii), the two outputs are determined by the precommitted variables, R_1 and R_2, since these variables determine the costs of production in stage (iii). That is,

$$X_1 = X_1(R_1,R_2) \tag{1.1}$$

$$X_2 = X_2(R_1,R_2) \ . \tag{1.2}$$

To impose rational expectations, we must suppose that both firms understand the dependence of X_1 and X_2 on R_1 and R_2. That is, both firms know equations (1.1) and (1.2). The profit of firm 2 can be written as

$$V_2(X_1,X_2,R_1,R_2) \equiv F_2(X_1,X_2) - C(X_2,R_2) - R_2 \ , \tag{2}$$

where $F_2(X_1,X_2)$ is total revenue accruing to firm 2. In stage (ii), R_1, already determined, is regarded by firm 2 as fixed; we suppose that firm 2 knows both $X_1(R_1,R_2)$ and $X_2(R_1,R_2)$. Assuming that V_2 is well behaved (i.e., we have an interior solution for R_2), firm 2's choice of R_2 satisfies:

$$\partial V_2/\partial R_2 = 0$$

or

$$[\partial F_2/\partial X_2 - \partial C(X_2,R_2)/\partial X_2]dX_2/dR_2 + \partial F_2/\partial X_1 \cdot dX_1/dR_2$$

$$- [\partial C(X_2,R_2)/\partial R_2 + 1] = 0 \quad . \tag{3}$$

The first term captures the marginal change in V_2 resulting from the dependence of X_2 on R_2. The second term captures the marginal change in V_2 resulting from the dependence of X_1 on R_2, and the third term captures the direct effect on V_2 of changing R_2.

For convenience, we assume that the market game (stage (iii)) is a Nash game in quantities. In this case, the first term in square brackets in (3) is zero, and firm 2's choice of R_2 satisfies

$$\partial F_2/\partial X_1 \cdot dX_1/dR_2 = [\partial C(X_2,R_2)/\partial R_2 + 1] \quad . \tag{4}$$

For the conventional case, $\partial F_2/\partial X_1$, $\partial X_1/\partial R_2 < 0$ so that the left-hand side of (4) is positive. Consider the non-strategic condition for R_2: Invest in R_2 until the reduction in total costs from more R&D (marginal benefit of R&D) just equals one (marginal costs of R&D) i.e., $-\partial C(X_2,R_2)/\partial R_2 = 1$. At this level of R&D, (4) would <u>not</u> be satisfied i.e., the conventional level of R&D is not profit maximizing in strategic settings. Rather, R&D must be further increased until (4) is satisfied. Given the quantity produced by firm 2 in the market game, R_2 will exceed the cost-minimizing value. The reason is that firm 2 can affect firm 1's quantity in the market game (which in turn affects firm 2's revenues) by manipulating R_2 since the market game is played subsequent to the choice of R_2. Thus, R_2 is of value to firm 2 for two distinct reasons: it is cost reducing (direct

effect) and it influences the quantity produced by firm 1 in the market game (indirect effect).

We can write firm 1's profit as:

$$V_1(X_1,X_2,R_1,R_2) \equiv F_1(X_1,X_2) - C(X_1,R_1) - R_1 \ , \qquad (5)$$

where $F_1(X_1,X_2)$ is firm 1's revenues in the market game. To impose rational expectations we suppose that firm 1 understands both the market game behavior shown in (1) and the dependence if R_2 on R_1 characterized by (4). We summarize (4) as $R_2 = R_2(R_1)$. To keep the mildly complicated chain of dependence straight, rewrite (5) as

$$V_1(R_1) \equiv F_1[X_1(R_1,R_2(R_1)) \ , \ X_2(R_1,R_2(R_1))]$$

$$- C[X_1(R_1,R_2(R_1)),R_1] - R_1 \ . \qquad (6)$$

Firm 1's choice of R_1 , (again assuming that the problem is well-behaved) satisfies:

$$\partial V_1/\partial R_1 = [\partial F_1/\partial X_1 - \partial C/\partial X_1][\partial X_1/\partial R_1 + \partial X_1/\partial R_2 \cdot dR_2/dR_1]$$

$$+ \partial F_1/\partial X_2 \cdot [\partial X_2/\partial R_1 + \partial X_2/\partial R_2 \cdot dR_2/dR_1]$$

$$- [\partial C/\partial R_1 + 1] = 0 \quad . \qquad (7)$$

We maintain the assumption that the market game is Nash in quantities. In this case, $\partial F_1/\partial X_1 = \partial C/\partial X_1$ so that (7) becomes

$$\partial F_1/\partial X_2 \ [dX_1/dR_1 + \partial X_2/\partial R_2 = dR_2/dR_1]$$

$$= [\partial C(X_1,R_1)/\partial X_1 + 1] \quad . \qquad (8)$$

Compare (4) and (8). For the conventional case, $\partial X_2/\partial R_2 > 0$ and $dR_2/dR_1 < 0$. Consequently, firm 1's incentives to "overinvest" in R&D are even stronger than firm 2's. Inspection reveals that the mechanism that led firm 2 to overinvest operates for firm 1 as well - the first term in the square brackets on the left-hand side of (8). In addition, as the first mover, firm 1 has the ability to manipulate firm 2's choice of R_2 which feeds back on firm 1's profits via the dependence of X_2 on R_2 and, in turn, F_1 on X_2.

It is then clear that in this case, the solution to the R&D game has both firms overinvesting in R&D in the sense that output produced is not produced at minimum cost, and, more importantly, the solution is asymmetric: $R_1 > R_2$, $X_1 > X_2$, $V_1 > V_2$.

This R&D example is particularly useful as a device for understanding the logical appeal of the paradigm outlined in Section I. R&D must precede the production of the output which makes use of the R&D; this sequential relationship is transparent to everyone concerned. This, of course, dictates that the entire game is a sequential game and, we assert, must be analyzed as such. We have imposed a particular structure on the sequential game by assuming that there is a first mover. This may be unappealing or inaccurate, but, were we to drop it, we would still be left with a sequential game -- a game in which firm 1 and firm 2 choose R_1 and R_2 simultaneously prior to the market game i.e., because R&D logically precedes the use in production of the fruits of R&D, the structure of the R&D game is sequential. (This is the game analyzed by Brander and Spencer (1982).)

In the R&D game considered here the incentives to manipulate the market game are driven through the cost side of the model--more R&D implies lower costs which imply a larger market share in the market game. Similar

incentives can arise through the demand side of a model. To illustrate this possibility, consider a sequential model in which advertising is the committed variable.

When advertising is the committed variable, revenues depend on the expenditure on advertising of the two firms, A_1 and A_2, as well as the quantities of the two firms, X_1 and X_2 : $F_i(X_1,X_2,A_1,A_2)$, $i = 1,2$. We assume that $\partial F_i/\partial A_i > 0$, and $\partial^2 F_i/\partial A_i \partial A < 0$: Total revenue is an increasing function of a firm's own expenditure on advertising and a firm's own marginal revenue of advertising is a decreasing function of expenditure by the other firm on advertising. We again examine a three stage game: (i) firm 1 chooses A_1, (ii) firm 2 subsequently choose A_2, taking A_1 as given and (iii) the two firms play a market game in which A_1 and A_2 are fixed. We again assume that the market game is a Nash game in quantities.

In the market game, X_1 and X_2 are determined by the precommitted variables A_1 and A_2 :

$$X_1 = X_1(A_1,A_2) \tag{9.1}$$

$$X_2 = X_2(A_1,A_2) \ . \tag{9.2}$$

The profits of firm 2 are

$$V_2(X_1,X_2,A_1,A_2) \equiv F_2(X_1,X_2,A_1,A_2) - C(X_2) - A_2 \ , \tag{10}$$

where $C(X_2)$ is the cost function. Taking A_1 as given, and knowing the dependence in equations (9), firm 2's choice of A_2 can be characterized by:

$$\partial V_2/\partial A_2 = [\partial F_2/\partial X_2 - C'] \cdot dX_2/dA_2$$

$$+ \partial F_2/\partial X_1 \cdot dX_1/dA_2$$

$$+ [\partial F_2/\partial A_2 - 1] = 0 \qquad (11)$$

(Again, we assume that V_2 is well-behaved and solutions are interior.) The first term on the right-hand side is zero as the market game is Nash in quantities. Thus, firm 2's choice of A_2 is described by:

$$\partial F_2/\partial X_1 \cdot dX_1/dA_2 = \partial F_2/\partial A_2 - 1 \quad . \qquad (12)$$

For the conventional case, $\partial F_2/\partial X_1$, $dX_1/dA_2 < 0$, so the left-hand side of (12) is positive. If firm 2 were unaware of the dependence in stage (iii) of X_1 on A_2 , then firm 2 would choose A_2 so that the direct effect of A_2 on revenues, $\partial F_2/\partial A_2$, is just balanced by the marginal cost of advertising, which is 1. In that event, the right-hand side of (12) would be zero. From (12), the recognition of the sequential structure of the advertising game leads firm 2 to choose more advertising than otherwise (given X_1).

Equation (12) obviously determines A_2 as a function of A_1, $A_2(A_1)$. Assuming that firm 1 understands this dependence and the dependence described in equations (9.1) and (9.2), we can write firm 1's profit as a function of A_1 .

$$V_1(A_1) = F_1[X_1(A_1,A_2(A_1)) \ , \ X_2(A_1,A_2(A_1)) \ , \ A_1,A_2(A_1)]$$

$$- C[X_1(A_1,A_2(A_1))] - A_1 \ . \qquad (13)$$

Firm 1's profit-maximizing choice of A_1 satisfies

$$\partial V_1/\partial A_1 = 0$$

or

$$[\partial F_1/\partial X_1 - C'][\partial X_1/\partial A_1 + \partial X_1/\partial A_2 \cdot dA_2/dA_1]$$

$$+ \partial F_1/\partial X_2 \cdot [\partial X_2/\partial A_1 + \partial X_2/\partial A_2 \cdot dA_2/dA_1]$$

$$+ \partial F_1/\partial A_2 \cdot dA_2/dA_1 + [\partial F_1/\partial A_1 - 1] = 0 \qquad (14)$$

Again, the first term is zero since the stage (iii) market game is Nash in quantities. Thus, firm 1's choice of A_1 is described by:

$$\partial F_1/\partial X_2 \cdot [\partial X_2/\partial A_1 + \partial X_2/\partial A_2 \cdot dA_2/dA_1]$$

$$+ \partial F_1/\partial A_2 \cdot dA_2/dA_1 = [1 - \partial F_1/\partial A_1] \quad . \qquad (15)$$

If firm 1 ignores the sequential nature of the game, A_1 would be chosen so that $1 = \partial F_1/\partial A_1$ i.e., the right-hand side of (15) would be zero. The three terms on the left-hand side in (15) describe firm 1's response to the sequential causality in the three stage game: A_1 directly reduces X_2 which increases firm 1's revenue; A_1 reduces A_2, which in turn reduces X_2, which then increases F_1; A_1 reduces A_2 which directly increases F_1. For the conventional case, the left-hand side of (15) is positive and the sequential structure of the game causes firm 1 to choose more advertising than otherwise (given X_1).

In the first R&D example, the chain of causality in the sequential game is driven by the effect of the committed variable (R&D) on a firm's own costs of production. In the second advertising example, the chain of causality is driven by the effect of the committed variable (advertising) on both a firm's own demand and the rival's demand. These observations suggest a useful way of constructing a taxonomy of strategic effects. First, what precommitted variables affect only a firm's own costs? Clearly research and development (Brander-Spencer (1982)), learning by doing (Spence (1981)), and expenditure on specific capital (Dixit (1980); Eaton-Lipsey (1981); Spence (1977)) qualify. What precommitted variables affect, in addition, the costs of rival firms? One important example arises where there exists monopsony with respect to some input. This possibility has not been carefully analyzed in the literature, but as we saw in Chapter 1, it is crucial to the banking industry. What precommitted variables affect demand conditions? Clearly advertising (Cubbin (1981)) and, in address models of monopolistic completion, the location of plants or brands in the appropriate space (Eaton-Lipsey (1979); Schmalensee (1978)). The approach developed in these address models has direct applicability to the banking sector to the extent that local branches have monopoly power in the placing of loans, and indirect applicability since the monopsony power inherent in branch banking is analogous to the monopoly power inherent in the address models of monopolistic competition.

REFERENCES

Ali, M. and Greenbaum, S. (1977). "A Spatial Model of the Banking Industry", Journal of Finance 32(4): 1283-1303.

Archibald, G.C. and Rosenbluth, G. (1975). "The "New" Theory of Consumer Demand and Monopolistic Competition", Quarterly Journal of Economics 80, 569-590.

Archibald, G.C., Eaton, B.C., and Lipsey, R.G. (forthcoming). "Address Models of Value Theory" in Mathewson-Stiglitz (eds.) New Developments in the Theory of Markets, MacMillan Co. Ltd.

Baltensperger, E. (1980). "Alternative Approaches to the Banking Firm", Journal of Monetary Economics 6: 1-37.

Barzel, Y. (1968). "Optimal Timing of Innovations", Review of Economics and Statistics 50(3): 348-355.

Baumol, W.J. (1952). "The Transactions Demand for Cash: An Inventory Theoretic Approach", Quarterly Journal of Economics 66: 545-56.

Baxter, W., Cootner, R., and Scott, P. (1977). Retail Banking in the Electronic Age, Allanheld, Osman & Co., Montclain, New Jersey.

Benston, G. (1972). "Economies of Scale of Financial Institutions", Journal of Money, Credit and Banking 4: 312-41.

Benston, G. and Smith, C., Jr. (1976). "A Transactions Cost Approach to the Theory of Financial Intermediation", Journal of Finance 31: 215-31.

Benston, G., Hanweck, G. and Humphrey, D. (1982). "Scale Economies in Banking", Journal of Money, Credit and Banking 14(4): 435-456.

Bond, D. and Shearer, R. (1972). The Economics of the Canadian Financial System: Theory, Policy and Institutions, Prentice Hall (Canada) Ltd., Scarborough, Ontario.

Brander, J.A. and Spencer, B.J. (1982). "Strategic Committment with R&D: The Symmetric Case", forthcoming in The Bell Journal of Economics.

Cubbin, J. (1982). "Advertising and the Theory of Entry Barriers", Economica 48: 289-298.

Demsetz, H. (1968). "Why Regulate Utilities", Journal of Law and Economics 11: 55-65.

Dixit, A. (1980). "The Role of Investment in Entry-Deterrence", The Economic Journal 90, 95-106.

Dixit, A. (1982). "Recent Developments in Oligopoly Theory", American Economic Review 72/21: 12-17.

Dixit, A. and Stern, N. (1982). "Oligopoly and Welfares", European Economic Review 19: 128-143.

Dudley, W. (1982). "Limit Pricing with An Application to Entry in Banking", unpublished Ph.D. dissertation, Department of Economics, University of California (Berkeley).

Eaton, B.C. (1976). "Free Entry in One Dimensional Models: Pure Profits and Multiple Equilibria", Journal of Regional Science 16, 21-33.

Eaton, B.C. and Lipsey, R.G. (1979). "The Theory of Market Preemption: The Persistence of Excess Capacity and Monopoly in Growing Spatial Markets", Economica 46, 149-58.

Eaton, B.C. and Lipsey, R.G. (1981). "Exit Barriers are Entry Barriers: The Durability of Capital as a Barrier to Entry", The Bell Journal of Economics 11, 721-29.

Eaton, B.C. and Wooders, M. (1983). "Sophisticated Entry in a Address Model of Monopolistic Competition", unpublished manuscript. University of Toronto.

Ederington, L. and Skogstad, S. (1977). "Measurement of Banking Competition", Journal of Money, Credit and Banking 9: 469-482.

Flannery, M. and Jaffee, D. (1975). The Economic Implications of an Electronic Monetary Transfer System, D.C. Heath and Company Ltd.

Friedman, M. (1955). "Comment" in (NBER) Business Concentration and Pricing Policy.

Graddy, D. and Kyle, R. (1978). "The Simultaneity of Bank Decision-Making; Market Structure and Bank Performance", Journal of Finance 34(1): 1-18.

McCall, A. and Peterson, M. (1977). "The Impact of De Novo Commercial Bank Entry", Journal of Finance 32(5): 1587-1604.

Prescott, E.C. and Visscher, M. (1977). "Sequential Location Among Firms with Foresight", The Bell Journal of Economics 8, 378-393.

Salop, S. (1979). "Strategic Entry Deterrence", American Economic Review 69, 335-38.

Schmalensee, R. (1978). "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry", The Bell Journal of Economics 9(2): 305-327.

Schmalensee, R. (1982). "The New Industrial Organization and the Economic Analysis of Modern Markets", in W. Hildenbrand (ed.), Advances in Economic Theory, Cambridge, Cambridge University Press.

Spence, M. (1977). "Entry, Capacity, Investment and Oligopolistic Pricing", The Bell Journal of Economics 8(2): 534-544.

Spence, M. (1981). "The Learning Curve and Competition", The Bell Journal of Economics 12(1): 49-70.

Stigler, G. (1968). "Price and Non-price Competition" Journal of Political Economy 22(1): 149-154.

Tobin, J. (1956). "The Interest Elasticity of Transactions Demand for Cash", Review of Economics and Statistics 38: 241-47.

Walker, D. (1978). "Economies of Scale and Electronic Funds Transfer Systems" Journal of Banking and Finance 2: 65-78.

_____ (1979). "An Analysis of Cash Dispenser and Automated Teller Activity Levels and Cost in the U.S.", Journal of Bank Research 7(9): 266-295.

Williamson, O. (1968). "Economies as an Antitrust Defence", American Economic Review 68(1): 112-123.